A new nonparametric copula framework for the joint analysis of river temperature and low flow characteristics for aquatic habitat risk assessment

SHAHID LATIF[1], Taha BMJ Ouarda[2], André- St-Hilaire[1], Zina Souaissi[3], and Shaik Rehana[4]

[1]Institut National de la Recherche Scientifique
[2]INRS-ETE, National Institute of Scientific Research
[3]Université du Québec à Montréal
[4]International Institute of Information Technology, Hyderabad

July 23, 2023

Abstract

The joint probability analysis of river water temperature (RWT) and low flow (LF) characteristics is essential as their combined effect can negatively affect aquatic species, e.g., ectotherm fish. Traditional multivariate models may not be as effective as copula-based methodologies. This study introduces a new multivariate approach, the nonparametric copula density framework, free from any distribution assumption in their univariate margins and copula joint density. The proposed framework utilized RWT and LF datasets collected at five different river stations in Switzerland. The study evaluates a nonparametric Gaussian kernel with six bandwidth selectors to model marginal densities. It employs nonparametric-based Beta kernel density, Bernstein estimator, and Transformation kernel estimator to approximate copula density with nonparametric and parametric margins. The performance of some parametric copula densities was also compared. The most justifiable models were employed to estimate bivariate joint exceedance probabilities and return periods (RPs). The Beta kernel copula with Gaussian kernel margins outperformed other models for most stations; Bernstein and Transformation copula with Gaussian kernel margins were better for only one station each. The univariate RPs (RWT or LF) are lower than the AND-joint but higher than OR joint case. As the percentile value of LF events (serve as a conditioning variable) increases, the bivariate joint RPs of RWT also increase. Higher values in RWT events result in higher RPs than lower values at the fixed percentile value of LF. All such estimated risk statistics are beneficial to analyze their mutual risk in aquatic habitats and freshwater ecosystems.

and-low-flow-characteristics-for-aquatic-habitat-risk-assessment

A new nonparametric copula framework for the joint analysis of river temperature and low flow characteristics for aquatic habitat risk assessment

Shahid Latif[1*], Taha B.M.J Ouarda[1], André- St-Hilaire[1, 2], Zina Souaissi[3], Shaik Rehana[4]

[1]Canada Research Chair in Statistical Hydro-Climatology, Institut national de la recherche scientifique, Centre Eau Terre Environnement, INRS-ETE, 490 De la Couronne, Québec City, QC, G1K 9A9, Canada

[2]Canadian Rivers Institute, University of New Brunswick, Canada

[3]Département des Sciences de la Terre et de l'atmosphère, Université du Québec à Montréal, Pavillon Président-Kennedy, Montréal, QC H2X 3Y7, Canada

[4]International Institute of Information Technology, Hyderabad, 500032, India

*Corresponding Author Email: Md_Shahid.LATIF@inrs.ca

Abstract:

The joint probability analysis of river water temperature (RWT) and low flow (LF) characteristics is essential as their combined effect can negatively affect aquatic species, e.g., ectotherm fish. Traditional multivariate models may not be as effective as copula-based methodologies. This study introduces a new multivariate approach, the nonparametric copula density framework, free from any distribution assumption in their univariate margins and copula joint density. The proposed framework utilized RWT and LF datasets collected at five different river stations in Switzerland. The study evaluates a nonparametric Gaussian kernel with six bandwidth selectors to model marginal densities. It employs nonparametric-based Beta kernel density, Bernstein estimator, and Transformation kernel estimator to approximate copula density with nonparametric and parametric margins. The performance of some parametric copula densities was also compared. The most justifiable models were employed to estimate bivariate joint exceedance probabilities and return periods (RPs). The Beta kernel copula with Gaussian kernel margins outperformed other models for most stations; Bernstein and Transformation copula with Gaussian kernel margins were better for only one station each. The univariate RPs (RWT or LF) are lower than the AND-joint but higher than OR joint case. As the percentile value of LF events (serve as a conditioning variable) increases, the bivariate joint RPs of RWT also increase. Higher values in RWT events result in higher RPs than lower values at the fixed percentile value of LF. All such estimated risk statistics are beneficial to analyze their mutual risk in aquatic habitats and freshwater ecosystems.

Keywords:

River water temperature, Low flow, Copulas, Beta kernel density, Bernstein estimator, Transformation kernel estimator, Joint return period, Conditional return period

1. Introduction

Compound events (or CE) can be defined by integrating the joint impact of two or more extreme or non-extreme events that co-occur or are in close succession, which might be a few hours to days apart (Seneviratne et al., 2012; Moftakhari et al., 2016; Zscheischler et al., 2018; Hendary et al., 2019). Such events might not be severe when defined individually, but when their effects are combined, they may be harmful. Interdependency between these events may be due to a common forcing mechanism driving the selected variables; thus, ignoring their collective impact or joint dependency would be risky. For instance, a rise in the river's temperature reduces the concentration of dissolved oxygen (Ficklin et al., 2013), and/or it can co-occur with low flows and doubly impact flora and fauna. Most water species (or aquatic organisms) have a specific temperature tolerance range. For example, there is a higher risk of proliferative kidney disease (PKD) in the brown trout population when temperatures rise above 15℃ (Strepparava et al., 2017). Vitellogenin (Vtg) concentration in brown trout's plasma, an indicator of estrogenic exposure, varies as a function of river temperature and the associated changes in estrogen uptake may have impacts on reproduction and development (above 19℃) (Korner et al., 2008). High water temperatures can increase fish mortality or limit their resources (Elliott & Hurley, 2001; Lund et al., 2002; Caissie et al., 2007). Besides this, the low flow period (or water scarcity) can be ecologically stressful; it can harm fish habitats and marine life (Daigle et al., 2011) and may reduce habitat connectivity (Fullerton et al., 2010). Also, this river flow reduction can increase river water temperature (Sinokrot and Gulliver, 2000; Humphries and Baldwin, 2003; Booker and Whitehead, 2021). Several pieces of literature in the past, for instance, St-Hilaire et al., 2011; Joshi et al., 2016; Lee et al., 2017; Ouarda et al., 2018; Caissie et al., 2019; Alobaidi et al., 2021; Souaissi et al., 2021; Ouarda et al., 2022; Abidi et al., 2022, performed univariate probability analysis of either extreme river temperature or low flow characteristics. Because of the negative correlation, a multivariate joint probability density function (pdf), joint cumulative distribution function (cdf) and their associated joint exceedance probabilities (or return periods) can better describe their compound effect. Their joint probability of occurrence (or exceedance probabilities) can be different than considering univariate probability distribution or frequency analysis (FA) of river water temperature or low flow characteristics.

The Modelling of CE usually considers the number of joint extremes or correlation structures between the variable of interest and highlights different extreme models (e.g.., Cloes and Twan, 1994; Coles et al., 1999; Coles, 2001; Samuels and Burt, 2002; Heffernan and Tawn 2004; Sevensson and Jones 2004; Boldi and Davidson 2007 and references therein). Recently, the copula function has been recognized as a highly flexible multivariate joint distribution tool frequently used in the Modelling of several hydrometeorological extremes (Joe, 1997; Nelsen, 2006; Shiau 2006; Salvadori and De Michele 2010; Salvadori et al., 2011; Vernieuwe et al., 2015; Latif and Mustafa 2020a; Chebana and Ouarda, 2021; Latif and Mustafa 2021; Latif and Simonovic 2022a, 2022b, 2022c and references therein). These studies frequently adopted parametric copula distribution settings, for instance, prior subjective assumptions about their univariate marginal probability density function (or PDF) and parametric class copulas (i.e., Archimedean, Elliptical or Extreme-value class etc.) in the joint pdf modelling. The parametric model incorporated in modelling univariate marginal density assumes that the random samples come from known populations whose pdf is predefined. No existing literature supports the selection of a fixed or specified distribution. Ideally, the random samples that typically follow different distributions need to be modelled separately without imposing any selected or fixed pdf a priori. In this regard, the data-driven model-based nonparametric kernel density estimations (KDE) can provide a bonafide distribution instead of a parametric density function and is free from any assumption (Silverman 1986; Adamowski 1989; Wand and Jones 1995; Kim et al. 2006, and references therein).

Classic parametric-based joint density framework incorporates conventional parametric models (i.e., bivariate normal or Gumbel etc.) (Goel 1998; Yue et al. 1999); or parametric copula functions (Nelsen 1996; Joe 1997). The question of how precisely and accurately the selected parametric copulas (and parametric marginal density) approximate the joint dependence structure between variables of interest, i.e., river water temperature (RWT) and corresponding low flow (LF), can be raised. Earlier studies, for instance, Genest and Rivest (1995), Shih and Louis (1995), and Bouezmarni and Rombouts (2008) highlighted a combination of the nonparametric marginal density with parametric copula density, called semiparametric copula settings. A few studies incorporated this semiparametric framework in hydrometeorological case studies (Karmakar and Simonovic 2009, Rauf and Zeephongsekul 2014; Latif and Mustafa 2021). However, the semiparametric approach still includes the involvement of parametric copulas, which might still be problematic, resulting in underestimating the actual multivariate joint density, as already discussed by Charpentier et al. (2006) and Rauf and Zeephongsekul (2014). Some statistical challenges exist, for instance, (1) parametric copulas dependence parameter estimation is quite time-consuming, (2) it would have spurious inferences if the hypothesis of fitted parametric copulas is violated, (3) it could demand extra precaution when selecting a suitable copula density for the given historical observations because different copulas capture the joint correlation structure differently, for instance, for Archimedean copula class, dependence parameter is restricted to some ' 'Kendall's tau τ_θ range. Such as, AMH copula is applicable to Kendall's tau τ_θ [-0.181, 0.333], or Clayton or extreme value class Gumbel copula, is only valid for positive dependency measures.

The nonparametric copula density estimation offers flexible alternatives and can adapt any joint mutual dependence structure without considering any specific probability density form for copulas and their univariate marginal density. Some earlier works include, for instance, Deheuvels and Hominal (1979) (nonparametric via empirical copula with nonparametric marginals), Gijbels and Mielniczuk (1990) (smooth kernel-based 2-D copula simulations via reflection method), Chen and Huang (2007) (bivariate kernel copula density via local linear estimator) etc. Besides this, a few studies, such as Harrell and Davis (1982), Chen (1999) and Charpentier et al. (2006), highlighted a nonparametric framework called Beta kernel estimator in the copula-based multivariate joint density simulation for financial data. Few studies incorporated this model in hydrometeorological studies, for instance, for rainfall events (i.e., Rauf and Zeephongsekul (2014)), flood events (i.e., Latif and Mustafa (2020b)) and wind energy modelling (i.e., Han and Chu 2021; Liang et al., 2022). The Beta kernel copula estimator can produce a minimum variance during estimation and is free of boundary bias problems. Besides this, Sancetta and Satchell (2004), Pfeifer et al. (2009), and Dieres et al. (2012) highlighted the efficacy of another nonparametric copula density approximation, called the Bernstein estimator. Bernstein estimator-based multivariate simulation can facilitate a higher consistency rate without any boundary bias problem. It can better estimate underlying mutual dependence than empirical copula density (Kulpa 1999, Weiss and Scheffer 2012). The efficacy of the Bernstein estimator in economics and financial data analysis is referred to above citation but is rarely incorporated in joint distribution modelling of extreme hydrometeorological characteristics. Latif and Slobodan (2022c) first introduced the Bernstein copula estimator with Beta kernel copula density through a comparative assessment in the joint probability modelling of storm surge and rainfall events in the compound flood risk assessments. Besides the aforementioned density estimators, Geenenens et al. (2014) discussed transformation estimators based on classical bivariate kernel estimations in joint probability density modelling.

It is useful to establish a multivariate joint framework by compounding the impact of high river water temperature and low flows as a better tool for better aquatic species management. The novelty of this study was to provide a methodological contribution, introducing and testing the efficacy of different nonparametric copula joint frameworks through a comparative assessment in the bivariate analysis of the above variable of interest. These nonparametric densities are tested for five different river stations in Switzerland. The nonparametric Gaussian KDE model with six different bandwidth selectors algorithms is proposed and compared with parametric class models in characterizing the univariate marginal pdfs of targeted variables. The proposed bivariate joint framework employed in this study, for instance, (1) a nonparametric Bernstein copula estimator with best-fitted nonparametric KDE margins, (2) a nonparametric Beta kernel copula density with best-fitted KDE margins, (3) a nonparametric Transformation estimator with best-fitted KDE margins, (4) Bernstein copula with best-fitted parametric class margins (5) Beta kernel copula density with best-fitted parametric margins, (6) Transformation estimator with best-fitted parametric margins, (7) parametric class copula density with best-fitted parametric margins (8) parametric copula density with nonparametric KDE margins. The most justifiable bivariate framework selected for each station was employed in estimating joint cumulative density, their exceedance probabilities and associated return periods for river water temperature and corresponding low flow characteristics.

The organization of this manuscript is as follows: Section 2 presents the theoretical background of the nonparametric bivariate joint probabilistic framework via Beta kernel copula density, Bernstein copula estimator, and Transformation kernel estimator. This section also discussed nonparametric kernel density estimation via the Gaussian function with different bandwidth selector approaches in modelling the marginal density of the variable of interest. Section 3 presents the study area details and delineation of bivariate random observations. Section 4 provides the results and discussions. Lastly, Section 5 presents the research conclusions and future works.

2. Methods

2.1. Nonparametric bivariate joint probabilistic framework

Examining the stress of river water temperature (RWT) or low flow (LF) individually may result in underestimating risk. The concept of multivariate joint exceedance probability and their associated return periods can provide a better assessment of the risk when considering the impact of both events jointly. The univariate probability approach would be confusing if correlated random variables (RWT and LF) describe the underlying event of interest. The copula function allows the individual Modelling of univariate margins and their joint dependence structure; this allows higher flexibility to select the most justifiable density functions for each variable, not necessarily from the same distribution family (Joe 1997; Nelsen 2006). Section 1 already pointed out a few statistical challenges in the parametric copula joint density framework. This study examined the adequacy of the different nonparametric-based copula combined with nonparametric and parametric margins for the RWT and LF characteristics. Figure 1 illustrates the methodological workflow model adopted in this study. The present study tested the performance of different nonparametric joint density models and compared their performance with previously selected parametric-based models in the joint probability density functions (JPDFs) and joint cumulative distribution functions (JCDFs) (or joint non-exceedance probabilities) between RWT and LF events. In our previous study, Latif et al. (2023) recognized different parametric copulas with parametric marginal distribution for the same stations between the same variable of interest. Referring to Figure 1, the efficacy of the six different Gaussian-based nonparametric kernel density models was developed to approximate the univariate marginal density of RWT and LF events. We also compared them to the performance of the best-fitted parametric models from Latif et al. (2023) for both variables. The second modelling stage proposed different bivariate structures by introducing nonparametric copula density with 1-D kernel density estimation (KDE) and parametric class margins. Finally, each station's most parsimonious joint density structure was selected to estimate bivariate joint exceedance probabilities and their associated joint return periods. The joint exceedance probability or return periods of RWT events, conditional to LF events (for different percentile values), are also examined.

Insert Figure 1

2.1.1. Nonparametric copula density modelling

The copula function connects multivariate joint probability distribution to the univariate marginal probability distribution of multiple individual random variables (Nelsen 2006 and Joe 1997). Copula relaxes the restriction in selecting univariate marginal pdf and copula dependence individually in two stages. Thus, the chosen univariate and multivariate functions do not necessarily belong to the same family of distributions. According to the Saklar theorem (Saklar 1959), for any continuous bivariate distribution whose cumulative distribution function, CDF, is F_{XY}, there exists a unique function 'C', called copula, such that

$$F_{XY}(x,y) = C\big(F_X(x), F_Y(y)\big), \qquad \forall(x,y)\mathbb{R}^2 \qquad (1)$$

In equation (1), $F_X(x), F_Y(y)$ are the univariate marginal cumulative distribution functions (CDFs).

This study incorporated a nonparametric copula density framework in the joint correlation structure and Modelling between RWT and LF events. Because the nonparametric copula framework is a distribution-free-based multivariate joint analysis, it can offer better flexibility than parametric copula settings. Through a comparative assessment, the adequacy of different bivariate joint probability frameworks is introduced.

2.1.2. Nonparametric joint density via Beta kernel copula estimator

The earlier study of the nonparametric approach in the copula density, i.e., Behnen et al. (1985), indicated that it could suffer from boundary bias problems, especially for symmetric kernels. The different nonparametric approaches for the joint copula density are selected based on procedures described in previous studies, for instance, Mirror image modification (Schuster 1985), transformed kernels (Wand et al. 1991), and boundary kernels (Müller 1991). Beta kernel density is employed in this study to approximate the nonparametric copula density, cojoined with Kernel density margins and parametric class models separately. It is naturally free from boundary bias compared to other standard kernel estimators, as indicated by Chen (1999), Brown and Chem (1999), and Renault and Scaliett (2004). Also, this estimator is much more consistent when the actual density is unbounded at the boundary.

Mathematically, if $X_1, X_2 \ldots \ldots, . X_P$, are the univariate random samples with known contact support [0, 1], 1-D Beta kernel density is estimated by (Charpentier et al., 2006);

$$d_h(x) = \frac{1}{p}\sum_{j=1}^{p} K(X_j, \frac{x}{h}+1, \frac{1-x}{h}+1) \quad (2)$$

$$K(x, s, t) = \frac{x^s(1-x)^t}{B(s,t)}, x \in [0,1] \quad (3)$$

Where

$$B(s, t) = \frac{\Gamma(s+t)}{\Gamma(s)\Gamma(t)} \quad (4)$$

Also,

$$K(x, s, t) = \frac{x^s(1-x)^t}{B(s,t)} = \frac{\Gamma(s)\Gamma(t)x^s(1-x)^t}{\Gamma(s+t)} \quad (5)$$

where K(..., s and t) is the Beta density function with parameters s and t, and 'h' is the kernel bandwidth [refer to Equation (1)]. Using the product of Beta kernel densities, also called the Beta kernel copula, at a point (u, v) is defined as bivariate copula joint density

$$c_h(x, y) = \frac{1}{ph^2}\sum_{j=1}^{p} K(X_i, \frac{x}{h}+1, \frac{1-x}{h}+1) \times K(Y_i, \frac{y}{h}+1, \frac{1-y}{h}+1) \quad (6)$$

In the above Equations (1) and (5) 'h' is the bandwidth of the Beta kernel density function and is estimated by minimizing the AMISE (Asymptotic mean integrated square error) followed by the Rule of thumb (ROT) and is calculated by (Nagler 2014)

$$h = \left(\frac{1}{8\pi}\frac{\varsigma(c)}{\xi(c)}\right)^{1/3} n^{-1/3} \quad (7)$$

This approach considered the Frank copula as a reference copula and is indicated by 'c.'

2.1.3. Nonparametric joint density via Bernstein copula estimator

The adequacy of the Bernstein polynomial in the joint probability functions appeared in earlier studies, for instance, Vitale (1975) and Tenbuh (1994). The Bernstein copula estimator can lack boundary bias problems and provide a higher consistency rate with a better estimation of underlying joint dependence structure than empirical copula (Sancetta and Satchell 2004; Diers et al. 2012). Also, it can perform better for an asymmetric mutual dependence (Bouezmarni et al., 2013).

Mathematically, the z-degree Bernstein polynomial is estimated by

$$B(z, a, b) = \binom{z}{a} b^a (1-b)^{z-a}, \qquad a = 0,1,2, \ldots \ldots, z \in \mathbb{N}; 0 \le b \le 1 \quad (8)$$

Then, for the bivariate joint distribution case, the 2-D Bernstein copula density is estimated by.

$$c(x_1, x_2) := \sum_{a_1=0}^{d_1-1}\sum_{a_2=0}^{d_2-1} p(a_1, a_2)\prod_{j=1}^{2} d_i B(d_j - 1, a_j, x_j), \qquad (x_1, x_2) \in [0,1]^2$$

(9)

where.

$$p(a_1, a_2) := P(\cap_{I=1}^{2}\{X_i = a_j\}), (a_1, a_2) \in [0,1]^2$$

(10)

where $X = (X_1, X_2)$ is the bivariate random samples having uniform margins over $\mathrm{P_j} := \{0, 1, 2, \dots \dots \dots \dots, \mathrm{z_j}\}$ with grid size $\mathrm{d_j} \in \mathbb{N}$.

2.1.4. Nonparametric density via Transformation method

Charpentier et al., (2006) introduced this nonparametric approach to kernel copula density estimation based on the earlier work of Devroye and Gyorfi (1985). A standard kernel density estimator can estimate this nonparametric density framework. Mathematically, the derivation of the estimator for the bivariate copula density is calculated by.

$$\mathrm{c_n}(\mathrm{x_1}, \mathrm{x_2}) = \frac{\hat{\mathrm{f}}_{\mathrm{n}}(\,\Phi^{-1}(\mathrm{x_1}), \Phi^{-1}(\mathrm{x_2}))}{\phi(\Phi^{-1}(\mathrm{x_1}))(\Phi^{-1}(\mathrm{x_2}))}, \qquad (x, y) \in [1, 2]^2 \tag{11}$$

Where Φ is the standard Gaussian cumulative distribution function, CDF and ϕ is the first order derivative.

The transformation estimator allows for a bandwidth matrix set by a rule of thumb (ROT), estimated by a normal reference rule on the transformed domain (Nagler et al., 2014).

$$\mathrm{BW} = \mathrm{n}^{-1/6} \hat{\Sigma}_{\mathrm{Z}}^{0.5} \tag{12}$$

Where $\hat{\Sigma}_{\mathrm{Z}}$ = empirical covariance matrix of $\Phi^{-1}(\mathrm{U_j})$ and $\Phi^{-1}(\mathrm{V_j})$, $\mathrm{j} = 1, 2, \dots \dots \dots, \mathrm{n})$

2.2. Nonparametric fitting of the univariate marginal distribution

The nonparametric kernel density estimation (KDE) is free from any prior distributional assumption. The univariate KDE was employed in the analysis of the marginal probability density of annual maximum rive water temperature (AMRWT) and LF having the following statistical properties (Rosenblatt 1956; Silverman 1986; Adamowski 1989; Hardle 1991);

$$\int_{-\infty}^{+\infty} \mathrm{K(x)dx} = 1 \tag{13}$$

Where $\mathrm{K(x)}$ is the univariate kernel density function. The generalized equation for kernel function is given by;

$$\mathrm{K_h(x)} = \frac{1}{\mathrm{h}} \mathrm{K}\left(\frac{\mathrm{x}}{\mathrm{h}}\right) \tag{14}$$

Where 'h' is the kernel bandwidth which can control the smoothness and roughness level in the kernel function's shape. Over-smoothing or under-smoothing can bypass actual content or can result in irregular density functions.
By taking the mean of equation (14), the univariate kernel density estimator was estimated by

$$\widehat{\mathrm{f_h}}(\mathrm{x}) = \frac{1}{\mathrm{nt}} \sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{K_h}\left(\frac{\mathrm{x - X_i}}{\mathrm{h}}\right) \tag{15}$$

In equation (15), n is the sample size, K(·) is the kernel function. Different bandwidth selection algorithms are discussed in the literature, with each method providing different estimates (Jones et al. 1996; Sharma et al. 1998; Sheather and John 1991; Chen et al. 2015). Selecting an appropriate bandwidth estimation is essential to control the shape of kernel density. This study tested the efficacy of Gaussian KDE with six different bandwidth selector algorithms in individually selecting the best-fitted univariate marginal pdfs for each station's RWT and LF variable chosen separately. For instance, the rule-of-thumb (ROT) given by Silvermann (1986), the rule-of-thumb (ROT) by Scott (1992), solve-the-equation (STE) and direct-plug-in (DPI) by Sheather & Jones (1991), biased cross-validation (BCV) and unbiassed cross-validation (UCV) (Scott and Tereel 1987; Santhos and Srinivas 2014)). The Gaussian (or Normal) KDE is widely accepted and can be expressed as

$$K_h\left(\frac{x-X_i}{h}\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-X_i)^2}{2h^2}} \tag{16}$$

Both Scott (1992) 's ROT (i.e., $h = 1.06 * \sigma * n^{-1/5}$; σ is the standard deviation), and Silverman's (1986) 's ROT (i.e., $h = 0.9 * \min\left(\sigma, \frac{IQR}{1.35}\right) * n^{-1/5}$; IQR is the interquartile range), bandwidth estimators are straightforward to calculate, where the performance of the second bandwidth estimator is more robust than the first, as already observed from the previous study. The other nonparametric bandwidth selector, which is based on cross-validation methods, for instance, UCV and BCV, can also give a better estimate of multimodal distribution. Conversely, the Sheather and Jones data-based bandwidth estimators, for example, the DPI and STE approaches, are also much more promising and can perform well (Elisa and Cao 2008; Chen 2015). This approach utilized pilot estimation of derivatives to select bandwidth. Besides nonparametric KDE margins, this study compared the performance of selected 1-D parametric models which were chosen from our previous study (Latif et al., 2023), joined with nonparametric and parametric copulas. The vector of unknown statistical or distribution parameters of the fitted parametric class 1-D density function is estimated via maximum likelihood estimation (MLE).

2.3 Model compatibility investigation

The performance of fitted nonparametric models, both 1-D marginal pdfs and bivariate copula joint density functions, is examined by employing different goodness of fit (GOF) tests, for instance, RMSE (Root Mean Square Error), MSE (Mean Square Error), MAE (Mean Absolute Error), NSE (Nash-Sutcliffe efficiency), AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion), HQC (Hannan-Quinn Information Criterion) (Akaike 1974; Singh et al., 2004; Wilmot and Matsura 2005; Zhang and Singh 2007; Gupta et al., 2005; Schwarz 1978; Farrel and Stewart 2006; Bennett et al., 2013; Moriasi et al., 2007; Singh et al., 2004; Hannan et al., 1979). The fitted models' fitness consistencies or GOF statistics are examined by observing gaps and dispensaries between the theoretical and empirical non-exceedance probabilities or CDFs. The minimum value of the above GOF test statistics reveals a better fit (or satisfactory performance), except for the NSE test, where a test value closer to 1 indicates the fitted (or theoretical) model is much closer to the empirical or optimal model. The NSE test values are numerically defined within a range of $-\infty$ (indicates for inferior model performance) to 1 (means ideal performance). Also, when its values lie between 0 to 1, it can further indicate a good agreement between empirical and theoretical observations. On the other side, the AIC test statistics comprise the lack of fit of the model on the one hand and the model's unreliability on the other hand. All three information criteria statistics, AIC, BIC and HQC, usually highlighted the trade-off relationship between model uncertainty and a number of fitted parameters. The HQC criterion does not exhibit asymptotically efficient criteria but also indicates higher consistency levels than AIC and BIC criteria (Haggag 2014). Besides, other statistical metrics, such as RMSE, MSE and MAE test, usually defines error statistics in the units of constituents of interests. Such as, test values closer to zero must be indicated for optimal model performance. The MAE statistics can minimize bias towards the large event relative to RMSE statistics and can be considered a better approach than the latter (Willmott and Matsuura, 2005; Bennett et al., 2013). The RMSE test performs relatively better than the MAE test for normally distributed samples (Chai and Draxler 2014), whereas MAE can perform somewhat better for skewed or multimodal distributed models. Besides, the NSE test compares data and residual variance structure.

The efficacy of the fitted bivariate copula densities was analyzed analytically by employing some additional GOF test measures and the above-discussed fitness statistics. For instance, K-S (Kolmogorov-Smirnov), mNSE (modified Nash-Sutcliffe Efficiency), IA (Index of Agreement), R2 (Coefficient of Determinations) and PBIAS (Percent Bias) are evaluated for each fitted bivariate density model for each station (Sorooshian et al., 1993; Moriasi et al., 2007; Ouarda et al., 2015; Onyutha 2021; Hoshin et al., 2009; Krause et al., 2005; Willmott et al., 1985; Legates et al., 1999; Kim et al., 2016; Nash and Sutcliffe 1970 and references therein). PBIAS measures the mean tendency of the simulated values to be larger or smaller than their empirical ones, where their optimal value is 0.0 (the positive value indicates overestimation bias or negative value for underestimation bias). Its low magnitude value must indicate for accurate model simulation or better fit. Besides, the coefficient of determination R^2 (range from 0 to 1 (better fit)) gives the proportion of the variance of one variable that is predictable from other variables. Similarly, IA (values range from 0 (indicates no agreements at all) to 1 (means perfect match)) statistic is a standardized measure of the degree of model prediction error or ratio of the mean square error and the potential error. Besides, the empirical distribution function-based KS test statistics measures the largest vertices between empirical and theoretical observations. Minimum the value K-S statistics can indicate a better fit. It is considered one of the most practical and general nonparametric approaches for comparing two samples or in model performance evaluation.

3. Study area and delineation of bivariate random samples

The proposed nonparametric copula joint framework is modelled for different time series of river temperature from gauging stations in Switzerland (Figure 2). There is a variation in the average watershed elevation ranging from 502 m to 1833 m, with catchment areas varying from 314 km^2 to a maximum of 6299 km^2. The annual cycle of river flow is moderate, with a higher level of inter-annual variability, which depends upon regional precipitation patterns and snowmelt (Michel et al., 2020).

Insert Figure 2

This study uses the data extracted from previous studies (Souaissi et al., 2021; Latif et al., 2023). The annual maximum rive water temperature (AMRWT) series were extracted from the daily time series for the summer months (May 1 to October 31) provided by the Swiss Federal Office for the Environment (FOEN). Another variable, river discharge during the low flow (LF) period, was defined by selecting the discharge value at the same calendar date of the annual RWT. The sample sizes for all selected five stations varied from 36 to 53 years.

4. Results and Discussion

4.1. Modelling of nonparametric KDE marginals of AMRWT and LF

Selecting a qualified marginal probability distribution is often a mandatory pre-requisite step before introducing it into a multivariate joint distribution framework. In this regard, the performance of different candidate models is often compared via the GOF test because multiple models would fit the data equally; however, it usually gives different quantile estimates, especially in distribution tails. The recent study by Latif et al. (2023) confirmed that AMRWT and corresponding LF exhibit zero serial correlation (or autocorrelation) for all selected stations. Also, the homogeneity test results show that both variables for all selected stations are homogenous except for station 2044, where changes occurred within the time series of AMRWT. Also, the nonparametric-based Mann-Kendall (M-K) test confirms nonstationarity (i.e., a significant positive trend) for the annual RWT series at stations 2044 and 2084.

The 1-D Gaussian KDE with six different bandwidth selector algorithms were fitted to historical time series to model univariate marginal probability distributions of AMRWT and corresponding LF series (refer to section 2.2). Supplementary Table (ST1) lists station-wise estimated bandwidth using different estimators. The compatibility of fitted candidate nonparametric 1-D models is compared for all selected stations using the various analytical-based GOF test statistics (refer to section 2.3). The empirical non-exceedance probabilities were estimated using the Gringgorten-based position-plotting formulae (Gringorten 1963) and were compared further with the theoretical observations estimated from the fitted candidate models. sThe following quantitative results are summarised below (refer to Supplementary Tables (ST 2a-e)):

1. Gaussian KDE with Silvermann ROT bandwidth estimator model was identified as the most parsimonious model in describing the univariate marginal distribution of AMRWT data at stations 2044, 2084, 2415 and 2473. Such as the majority of fitness tests are in support of the selected model. For instance, at station 2044, the selected model exhibits the minimum GOF test value of RMSE (0.0213 < RMSE of other candidate models), MSE (0.0004 < MSE of other candidate models), MAE (0.0161 < MAE of other candidate models), AIC (-405.65 < AIC of other candidate models), BIC (-403.68 < BIC of other candidate models), HQC (-404.89 < HQC of other candidate models) and higher values of NSE (0.9945 > NSE of other candidate models) statistics). For station 2106, the Gaussian KDE with UCV is the best model for the same variable, AMRWT. For instance, RMSE (0.0231 < RMSE of other candidate models), MSE (0.0005 < MSE of other candidate models), MAE (0.0179 < MAE of other candidate models), NSE (0.9935 > NSE of other candidate models), AIC (-359.32 < AIC of other candidate models), BIC (-357.45 < BIC of other candidate models), HQC (-358.62 < HQC of other candidate models).
2. Similarly, for stations 2044, 2415 and 2473, the Gaussian KDE with ROT bandwidth selector model was the best fit for LF values. For instance, at station 2044, the selected density function exhibits a minimum value RMSE (0.0226< RMSE of other candidate models), MSE (0.0005 < MSE of other candidate models), MAE (0.0177 < MAE of other candidate models), AIC (-399.62 < AIC of other candidate models), BIC (-397.65 < BIC of other candidate models), HQC (-398.86 < HQC of other candidate models). It also exhibits a higher value of NSE (0.9938 > NSE of other candidate models). For Station 2084, it was the Gaussian KDE with UCV model; for Station 2106, the Gaussian KDE with STE model was selected best. For instance, at station 2106, RMSE($RMSE_{2106}$ (Gaussian KDE with STE model) < $RMSE_{2106}$ of other candidate models), MSE (MSE_{2106} (Gaussian KDE with STE model) < MSE_{2106} of other candidate models), MAE (MAE_{2106} (Gaussian KDE with STE model) < MAE_{2106} of other candidate models), AIC (AIC_{2106} (Gaussian KDE with STE model) < AIC_{2106} of other candidate models), BIC (BIC_{2106} (Gaussian KDE with STE model) < BIC_{2106} of other candidate models), HQC (HQC_{2106} (Gaussian KDE with STE model) < HQC_{2106} of other candidate models), and NSE (NSE_{2106} (Gaussian KDE with STE model) > NSE_{2106} of other candidate models) respectively.

The same tables also reveal that all selected KDE 1-D models outperformed the best-performing parametric distributions, which were fitted to both AMRWT and LF variables in our previous study for the same stations. For instance, the selected nonparametric Gaussian KDE with ROT model captures marginal behaviour for both variables at station 2044 much more effectively than the parametric-based Logistic-2P models (refer to ST 2a-e). For example, for AMRWT variable ($RMSE_{\text{GAUSSIAN ROT MODEL}}$ (0.0213) < $RMSE_{\text{PARAMETRIC LOGISTIC-2P MODEL}}$ (0.0222), $MSE_{\text{GAUSSIAN ROT MODEL}}$ (0.0004) < $MSE_{\text{PARAMETRIC LOGISTIC-2P MODEL}}$ (0.0005), $MAE_{\text{GAUSSIAN ROT MODEL}}$ (0.0161) < $MAE_{\text{PARAMETRIC LOGISTIC-2P MODEL}}$ (0.0181), $AIC_{\text{GAUSSIAN ROT MODEL}}$ (-405.65) < $AIC_{\text{PARAMETRIC LOGISTIC-2P MODEL}}$ (-399.33), $BIC_{\text{GAUSSIAN ROT MODEL}}$ (-403.68) < $BIC_{\text{PARAMETRIC LOGISTIC-2P MODEL}}$ (-395.39), $HQC_{\text{GAUSSIAN ROT MODEL}}$ (-404.89) < $HQC_{\text{PARAMETRIC LOGISTIC-2P MODEL}}$ (-397.81), $NSE_{\text{GAUSSIAN ROT MODEL}}$ (0.9945) > $NSE_{\text{PARAMETRIC LOGISTIC-2P MODEL}}$ (0.994)), and variable LF ($RMSE_{\text{GAUSSIAN ROT MODEL}}$ (0.0226) < $RMSE_{\text{PARAMETRIC LOGISTIC-2P MODEL}}$ (0.0282), $MSE_{\text{GAUSSIAN ROT MODEL}}$ (0.0005) < $MSE_{\text{PARAMETRIC LOGISTIC-2P MODEL}}$ (0.0008), $MAE_{\text{GAUSSIAN ROT MODEL}}$ (0.0177) < $MAE_{\text{PARAMETRIC LOGISTIC-2P MODEL}}$ (0.0247), $AIC_{\text{GAUSSIAN ROT MODEL}}$ (-399.62) < $AIC_{\text{PARAMETRIC LOGISTIC-2P MODEL}}$ (-374.03), $BIC_{\text{GAUSSIAN ROT MODEL}}$ (-397.65) < $BIC_{\text{PARAMETRIC LOGISTIC-2P MODEL}}$ (-370.09), $HQC_{\text{GAUSSIAN ROT MODEL}}$ (-398.86) < $HQC_{\text{PARAMETRIC LOGISTIC-2P MODEL}}$ (-372.52), $NSE_{\text{GAUSSIAN ROT MODEL}}$ (0.9938) > $NSE_{\text{PARAMETRIC LOGISTIC-2P MODEL}}$ (0.9903)). Supplementary Table ST3 summarises the estimated parameters of the best-fitted parametric distribution for each station via the maximum likelihood estimator (MLE). Visual inspection confirms the adequacy of fitted nonparametric 1-D models using overlapped probability density function (PDF) plots (refer to Supplementary Figures (SF 1-2)). It was found that all selected KDE models adequately captured the distribution behaviour of the targeted random variables and supported the analytical investigation. In conclusion, these estimated results confirmed our initial hypothesis about the robustness of nonparametric KDE pdf over the parametric distributions in modelling the marginal distribution of AMRWT and LF events.

4.2. Nonparametrically joint correlation modelling between AMRWT and LF events

Our recent study, Latif et al. (2023), confirmed, as expected, the existence of a negative correlation between LF and AMRWT events (Pearson, Kendall's and Spearman; significant at a 95% confidence interval). Graphically, 2-D scatter plots, Chi-plots, Kendall's or K-plot between the AMRWT and corresponding LF for the selected stations also confirmed the negative correlation. One of the most considerable statistical flexibilities of nonparametric copula density is that it can adapt to any type of mutual dependence structure to the given bivariate pairs without having any distributional assumption. The parametric copula models notoriously lack flexibility and can bear the risk of misspecification.

Obtained Framework 1: Nonparametric margins with nonparametric copula (NPMNPC) models:

The bivariate joint frameworks included the Beta kernel copula density, Bernstein copula estimator and Transformation estimator. All selected bivariate densities were joined with best-fitted 1-D nonparametric margins in the joint probability modelling between AMRWT and LF (refer to Equations 6, 9 and 11). The bandwidth of the fitted Beta kernel copula estimator is estimated by the Rule of thumb (ROT), followed by Equation 7 (refer to Method 2.1.1). In working with the Bernstein copula estimator, their coefficients are adjusted, followed by Weiss and Scheffer (2012). Also, the bandwidth matrix for the transformation kernel estimator was estimated by Equation (12). Tables 1(a-e) list the developed bivariate joint density models and the estimated bandwidth (only for Beta kernel copula and Transformation kernel density).

Obtained Framework 2: Parametric margins with parametric copula (PMPC) model:

This bivariate parametric joint framework introduces the best-fitted parametric class 2-D copulas selected from the previous study joined with best-fitted 1-D parametric marginal pdfs (refer to Supplementary Tables (ST 3, ST 4, and Tables 1(a-e)). For instance, rotated Clayton copula (90 degrees) (for station 2044), rotated BB8 copula (270 degrees) (for station 2084), rotated Joe copula (90 degrees) (for station 2106), rotated Tawn type-1 copula (90 degrees) (for station 2415), and rotated Clayton copula (90 degrees) (for station 2473). The copula dependence parameters were estimated using the maximum pseudo-likelihood (MPL) estimator (Latif et al., 2023).

Obtained Framework 3: Parametric margins with nonparametric copula (PMNPC) model:

These semiparametric-based bivariate frameworks employed the Beta kernel copula, Bernstein copula estimator and Transformation kernel estimator density individually with selected best-fitted parametric margins (refer to Table 1(a-e) and ST3).

Obtained Framework 4: Nonparametric margins with parametric copula (NPMPC) model:

These semiparametric bivariate frameworks introduce the best-fitted KDE margins of the AMRWT and corresponding LF together with the most parsimonious 2-D parametric copulas (refer to Tables 1(a-e) and ST4).

4.3. Model's performance comparison

The efficacy of all the above nonparametric, semiparametric and parametric bivariate copula joint density frameworks was analyzed and compared, for each station individually, based on different GOF test statistics (refer to Tables 1(a-e)). The empirical bivariate joint non-exceedance probabilities or CDFs were obtained from the empirical copula, followed by Deheuvels (1979). Different model fitness test statistics were employed; refer to section 2.3. The nonparametric Analytical investigation reveals that Beta kernel copula density with KDE (GAUSSIAN-Silverman Rule-of-thumb ROT) margins was found to be the most justifiable density and best performance in capturing joint correlation structure between AMRWT and LF at stations 2044 and 2106. The selected joint density is in favour due to the majority of estimated GOF test values in support compared to other developed bivariate models at this station. For instance, the estimated GOF measures of the selected model at station 2044 ($\text{K-S}_{\text{BETA GAUSS. ROT MODEL}} < \text{K-S}_{\text{OTHER MODELS}}$, $\text{RMSE}_{\text{BETA GAUSS. ROT MODEL}} < \text{RMSE}_{\text{OTHER MODELS}}$, MSE_{BETA}

$_{\text{GAUSS. ROT MODEL}}$ < $MSE_{\text{OTHER MODELS}}$, $(NSE\ \&\ mNSE)_{\text{BETA GAUSS. ROT MODEL}}$ > $(NSE\ \&\ mNSE)_{\text{OTHER MODELS}}$, $(AIC, BIC\ \&\ HQC)_{\text{BETA GAUSS. ROT MODEL}}$ < $(AIC, BIC\ \&\ HQC)_{\text{OTHER MODELS}}$, $(IA\ \&\ R^2)_{\text{BETA GAUSS. ROT MODEL}}$ > $(IA\ \&\ R^2)_{\text{OTHER MODELS}}$. Besides, the estimated PBIAS value for this model is closer to zero, indicating less risk of overestimation bias than other bivariate density models. Similarly, for station 2084, the Bernstein copula with KDE (GAUSSIAN) Silverman Rule-of-thumb (ROT)- KDE (GAUSSIAN) Unbiased cross-validation (UCV) Scott and Terrell (1987) margin was recognized as the best-fitted model for that station. The transformation estimator with KDE (GAUSSIAN) UCV- KDE(GAUSSIAN) STE margins describe the most justifiable bivariate dependence at station 2415. For station 2473, the Beta kernel copula with parametric-based Logistic-2P and Lognormal-2P margins outperformed other bivariate densities. Overall, the Beta kernel copula density outperformed and was much more efficient than the Bernstein copula estimator and Transformation estimator for most stations when it is used with KDE marginal densities. Conversely, the Beta kernel copula surpassed other nonparametric joint densities when joined with parametric class marginal density for station 2473. Besides the selected bivariate model for each station, for instance, the performance of the Transformation kernel estimator with Gaussian KDE margins outperformed the Bernstein estimator at stations 2044 and 2106, while its performance was inferior at stations 2084 and 2473. Also, when the nonparametric Beta kernel, Bernstein estimator and Transformation estimator functions were used with the best-fitted parametric margins, the performance was less robust and inferior to when the same copula density model was fitted with nonparametric margins for all stations, except for 2473 (Beta kernel copula with parametric class margins). For instance, at station 2044, refer to Table 1a, when comparing the estimated GOF statistics of fitted NPMNPC models versus PMNPC models, ($K\text{-}S_{NPMNPC}$ < $K\text{-}S_{PMNPC}$), ($RMSE_{NPMNPC}$ < $RMSE_{PMNPC}$), (MSE_{NPMNPC} < MSE_{PMNPC}), (MAE_{NPMNPC} < MAE_{PMNPC}), ($NSE\ \&\ mNSE_{NPMNPC}$ > $Mnse\ \&\ NSE_{PMNPC}$), ($AIC, BIC\ \&\ HQC_{NPMNPC}$ < $AIC, BIC\ \&\ HQC_{PMNPC}$), ($IA_{NPMNPC}$ > IA_{PMNPC}). Besides, the estimated PBIS statistic is closer to zero for NPMNPC models than PMNPC models.

It was also found that the best-fitted parametric copulas modelled with parametric marginals or PMPC models, selected from the previous study, had inferior performance or were less robust than nonparametric copulas' density or NPMNPC models for all selected stations (refer to same Tables 1a-e). However, when the same parametric copulas were used with the best-fitted nonparametric KDE margins or NPMPC models, their performances were much better than when parametric copulas were fitted to parametric margins or PMPC for all selected stations. For instance, at station 2044, when comparing the performance based on their estimated fitness measures between PMPC and NPMPC models, ($K\text{-}S_{NPMPC}$ < $K\text{-}S_{PMPC}$), ($RMSE_{NPMPC}$ < $RMSE_{PMPC}$), (MSE_{NPMPC} < MSE_{PMPC}), (MAE_{NPMPC} < MAE_{PMPC}), ($NSE\ \&\ mNSE_{NPMPC}$ > $Mnse\ \&\ NSE_{PMPC}$), ($AIC, BIC\ \&\ HQC_{NPMPC}$ < $AIC, BIC\ \&\ HQC_{PMPC}$), (IA_{NPMPC} > IA_{PMPC}) etc.

The performances of bivariate joint densities were examined graphically using an overlapped 2-D scatterplot between historical bivariate random pairs (indicated by the red colour) with a set of generated pairs (sample size, N=1000, indicated by the light-blue colour) estimated from fitted candidate bivariate densities. Refer to Supplementary Figures (SF 3(a-e)), it is illustrated that all the selected bivariate joint densities performed adequately; the generated random pairs (in light blue) overlapped with the natural mutual dependence of the historical samples (in red) for all selected stations. At this point, it is concluded that our initial hypothesis about the flexibility of nonparametric copula density in the joint dependency modelling exhibits is superior to fully parametric frameworks and is therefore deemed more suitable in the joint probability modelling of AMRWT and LF events. Supplementary Figure SF4 shows a 3-D scatterplot of the joint cumulative probability distributions, also called joint non-exceedance probabilities, derived from each station's best-fitted bivariate copula density using the historical observational events, AMRWT and corresponding LF. Besides, Supplementary Figures (SF 5a-e) and (SF 6) illustrate the surface density plots and contour plots of selected bivariate copula joint density fitted between AMRWT and LF events for each station.

Insert Tables 1 (a-e) here

4.4. Risk evaluation via joint and conditional distributions

4.4.1. Primary joint return period for OR and AND case

The derived nonparametric copula density framework further examined the joint probability distribution relationship between AMRWT and corresponding LF events that can stress certain endemic fish species when both variables occur concurrently. Supplementary Figures (SF 7a-b) illustrate the univariate return periods of AMRWT and LF events estimated from best-fitted 1-D marginal distributions. The univariate approach in the risk evaluation would result in an underestimation. Their joint correlation behaviour was captured based on multivariate joint and conditional joint exceedance probabilities and their associated joint and conditional return periods. Two different joint probability cases and return periods are considered in this study (Salvadori, 2004; Graler et al., 2013; Reddy and Ganguli, 2012; Salvadori et al., 2015), as summarised below:

1. When both annual maximum river water temperature (AMRWT) and concomitant low flow (LF) simultaneously exceed a specific threshold value (say amrwt and lf), i.e., $AM\mathrm{RWT} \geq \mathrm{amrwt\ AND\ LF} \geq$ lf), called the AND-joint case and the associate return periods are estimated by.

$$\mathrm{T}^{\mathrm{AND}}_{\mathrm{AMRWT,LF}} = \frac{1}{1-\mathrm{F(AMRWT)}-\mathrm{F(LF)}+\mathrm{C(AMRWT,\ \ LF)}} = \frac{1}{1-(\mathrm{F(AMRWT)}+\mathrm{F(LF)}-\mathrm{C(AMRWT,\ \ LF)})} \quad (17)$$

Where, $\mathrm{F(AMRWT)}$ and $\mathrm{F(LF)}$ are the best-fitted univariate marginal cdfs (or univariate non-exceedance probabilities); $\mathrm{C(AMRWT,\ \ LF)}$ is the joint cdf values which are estimated using the best-fitted bivariate copula joint density framework. Equation (17) examined the risk when AMRWT and LF events coincide (simultaneous occurrence). Also, the denominator term "$\mathrm{F(AMRWT)} + \mathrm{F(LF)} - \mathrm{C(AMRWT,\ \ LF)}$" define the joint cumulative probability distribution or joint non-exceedance values in the simultaneous occurrence of AMRWT and LF event (refer to Supplementary Figure SF8).

2. When AMRWT or LF exceed a specific threshold value (i.e., AMRWT ≥ amrwt OR LF ≥ lf), called the OR-joint case, the associated return periods are estimated.

$$\mathrm{T}^{\mathrm{OR}}_{\mathrm{AMRWT,LF}} = \frac{1}{1-\mathrm{C(AMRWT,\ \ LF)}} \quad (18)$$

Tables 2(a-e) compare each station's estimated univariate versus bivariate return periods. In this Table, at first, the univariate design variable quantiles for AMRWT and corresponding LF events are calculated using the quantiles functions (from best-fitted KDE marginal models) for specified annual exceedance probabilities (AEP) or univariate return periods, for instance, 2, 5, 10, 20, 30, 50, 79, and 100 years. The bivariate OR and AND-joint return periods are estimated for the different combinations of designed AMRWT and LF (refer to same Tables 2 (a-e) and Supplementary Figure (SF9)). It is found that AND-joint return periods are higher than OR-joint case for any combination of bivariate design events at any station, i.e., $\mathrm{T}^{\mathrm{AND}}_{\mathrm{AMRWT,LF}} > \mathrm{T}^{\mathrm{OR}}_{\mathrm{AMRWT,LF}}$. It further reveals that there is less chance (i.e., less probable) or frequency in the occurrence of bivariate events simultaneously in the AND-joint case than in the OR-joint case. Also, OR-joint return periods are less than univariate return periods estimated by considering univariate marginal CDF of the best-fitted model for AMRWT or LF events, i.e., $\mathrm{T}^{\mathrm{OR}}_{\mathrm{AMRWT,LF}} <$ ($\mathrm{T}^{\mathrm{UNIVARIATE}}_{\mathrm{AMRWT}}$ or $\mathrm{T}^{\mathrm{UNIVARIATE}}_{\mathrm{LF}}$). Also, the joint return periods were estimated using the historical bivariate events for both joint cases and illustrated using 3-D scatterplots (referring to Supplementary Figures (SF10a-e)). For instance, on June 5, 2019, at station 2044, the bivariate events, AMRWT and LF, were 20.26 ℃ and $34.975\ \mathrm{m^3/sec}$, having AND and OR-joint return period were 112.20 years and 1 year. Similarly, at station 2473, on August 11 2019, AMRWT and LF events were 19.74 ℃ and $129.168\ \mathrm{m^3/sec}$, the AND and OR joint return periods were 196.23 years and 1.29 years.

When the temperature rises above 19 °C, it can stop the growth of brown trout (*Salmo trutta)* or influence the concentration of vitellogenin (Vtg) in brown trout's plasma. From the Supplementary Figures (SF11 a-b), it is inferred that the estimated AMRWT events exceed this threshold at the higher exceedance probability (or when its return period is 2 years or above) for all stations except station 2473. For instance, at 2-year and 10-year return periods, the estimated AMRWT is 25.06 °C and 26.82°C (for station 2044), 18.56 °C and 20.89 °C (for station 2084), 21.65 °C and 23.48 °C (for station 2106), 24.65 °C and 25.61 °C (for station 2415). However, for station 2473, at the same return periods, the AMRWT quantiles are 16.53 °C and 17.98 °C, which attained its value above 19 °C (threshold) at 50 years or above return periods.

During the low water periods, the small wetted areas lead to a decrease in available physical habitat that can also harm fish. From the same Table 2 (a-e), the absolute and specific discharge values are compared for different stations at different return periods. From Supplementary Figures (SF12) and (SF13), it is found that station 2084 exhibits the highest value of specific discharge and station 2106 exhibits the lowest value as compared to other stations. For instance, at a return period of 30 years, the estimated LF was 25.99 m^3/sec (with specific discharge = 0.015204 ($\frac{m^3}{sec}/km^2$), with drainage surface area = 1709.42 m^2), 28.16 m^3/sec (with specific discharge = 0.089465 ($\frac{m^3}{sec}/km^2$),, drainage surface area = 314.76 m^2), 8.16 m^3/sec (with specific discharge = 0.015204 ($\frac{m^3}{sec}/km^2$), drainage surface area = 942.92 m^2), and 292.82 m^3/sec (with specific discharge = 0.046485 ($\frac{m^3}{sec}/km^2$),, drainage surface area = 6299.198 m^2) at stations 2044, 2084, 2106, and 2473, respectively. In conclusion, it is found that considering only a univariate case of return periods would be problematic; it can mislead the risk assessments when compounding their joint correlation behaviour and would result in the underestimation of risk.

Insert Tables 2 (a-e)

4.4.2. Joint return periods of AMRWT events conditional to LF events

The risk evaluation between AMRWT and LF events via conditional return periods relies on their conditional joint probability relationship. The joint return of AMRWT, given various percentile values of corresponding LF, are estimated using the best-fitted derived bivariate models for the case, $T_{A|B\leq b}$ (Shiau, 2006; Zhang and Singh, 2006; Salvadori and De Michele, 2010; Sraj et al., 2014) is estimated by

$$\mathrm{T}_{\mathrm{AMRWT|LF\leq lf}} = \frac{1}{(1-\mathrm{C(rwt,\ lf)/F(lf)})} \tag{19}$$

The conditional return periods of AMWRT were estimated considering different percentile values of LF events, for instance, 5th, 25th, 50th, 75th, 90th and 95th percentiles. Referring to Figures 3 (a-e), it is found that for all selected stations, return periods of AMRWT increase with an increase in percentile values of their corresponding LF events. For instance, at station 2044, on July 27, 1979, the AMRWT was 24.05 °C, the joint return period for the aforementioned percentiles was 1.01 years (when corresponding LF ≤ 8.0456 m^3/s (5th percentile)), 1.17 years (when corresponding LF ≤ 11.029 m^3/s (25th percentile)), 1.20 years (when LF ≤ 20.4584 m^3/s (90th percentile)), and 1.22 years (when LF ≤ 21.7814 m^3/s (95th percentile)). Similarly, at station 2473, the AWRWT was 16.21°C on August 11, 2019, and the joint return period was 1.24 years, 1.26 years, 1.46 years, 1.53 years when the corresponding LF ≤ value at the 5th percentile (123.43 m^3/s), LF ≤ value at 25th percentile (140.84 m^3/s), LF ≤ value at 75th percentile (217.53 m^3/s) and LF ≤ value at 90th percentile (257.71 m^3/s). Besides, for all selected stations, the higher bivariate return periods were obtained by fixing the percentile values with an increase in the value of AMRWT. For instance, at station 2106, by fixing the percentile value of the conditional variable, LF, say 75th (5.77 m^3/s), the return period of AMRWT was 1.13 years (when AMRWT was 20.94 °C on June 23, 2002), 22.44 years (when AMRWT was 24.4 °C on August 13, 2003), 1.54 years (when AMRWT was 21.55 °C on August 2, 2004), 3.10 years (when AMRWT was 22.43 °C on July 21, 2006), 10.01 years (when AMRWT was 23.75 °C on July 7, 2015).

Insert Figures 3(a-e)

5. Research Conclusions

Compounding the joint impact of river water temperature (RWT) and low flow (LF) events that can potentially harm the aquatic habitat provides more significant information to managers than just considering these events individually. Both variables exhibited a negative dependence structure already confirmed by a previous study (Latif et al., 2023). Thus, incorporating a multivariate joint probability distribution framework and the notations of joint and conditional joint return periods can comprehensively measure the risk associated with these bivariate events. The applicability of traditional multivariate parametric models or copulas in joint density modelling would have some statistical limits regarding prior distribution assumptions in their univariate margins and copula joint pdf, which also can lack flexibility. Also, if the underlying statical hypothesis is violated, it could lead to misspecification. This study provided a methodological contribution by incorporating a new approach via a nonparametric-based copula distribution approach in joint probability density modelling of AMRWT and LF for five different river locations in Switzerland. The nonparametric copula framework is free from any distribution assumption. It can adapt any joint mutual correlation structure without assuming any fixed or specific probability density form for either copula or univariate margins.

The joint probability analysis proposed different nonparametric copula densities through comparative assessments: Beta kernel density, Bernstein copula estimator, and Transformation kernel estimator. All such nonparametric copula densities were combined with nonparametric and parametric class marginal densities in establishing a joint dependence structure. The performance of parametric copulas joined with best-fitted nonparametric KDE and parametric class margins were also tested and compared in the joint dependence between AMRWT and LF events. The Gaussian KDE with six different bandwidth estimators were introduced and compared with parametric margins in modelling the univariate marginal pdfs of AMRWT and LF events. Model compatibility investigation confirmed that nonparametric KDE margins outperformed parametric class marginal density for all selected stations. Also, the ROT Silvermann bandwidth estimator with Gaussian KDE performed better for both variables at most stations than other models. Based on different GOF test statistics and also graphical visual inspection, a comprehensive model performance investigation confirmed that the nonparametric copula joint framework (i.e., nonparametric copulas with nonparametric margins) outperformed the other models for all selected stations. For instance, Beta kernel copula with Gaussian KDE-Silvermann ROT margins captured the mutual correlation between AMRWT and LF events in a better manner at station 2044. At station 2473, the same Beta copula density performed well when joined with parametric marginal densities than nonparametric Gaussian KDE margins. Overall, Beta kernel copula density satisfied the most justifiable for most of the stations.

The best-fitted bivariate joint density was selected for each station to estimate the joint return period for OR and AND joint cases. The return periods were observed for different combinations of designed RWT and LF events estimated at different univariate return periods. It is found that the chance of simultaneous occurrence of both AMRWT and LF events is lower or less in AND joint case compared to OR joint case for all selected stations. The accountability of only univariate return periods in the risk evaluation would result in underestimation when multiple random variables significantly impact when they occur jointly. The univariate return periods considering AMRWT or LF events are found to be less than the AND-joint case but higher than the OR joint case.

The derived nonparametric models estimated the conditional joint return period of AMRWT events given various percentile values of LF events for case $T\_(AMRWT|LF \leq lf)$, for all stations. It is found that the return periods of bivariate events increase with an increase in the percentile value of LF. It is also found that, for all stations, the higher value in AMRWT events will result in high bivariate return periods compared to a lower value at the fixed percentile value of LF (conditioning variable). In conclusion, the estimated return periods could provide insight into the relative mutual dependence behaviour of river thermal-low flow risk for aquatic species in Swiss rivers.

Acknowledgement

The authors thank the National Sciences and Engineering Research Council of Canada (NSERC) and the Canada Research Chair Program for their support in funding this research. Also, thanks to the Ministry of Science & Technology, Department of Science and Technology (DST), Government of India and India-Canada Centre for Innovative Multidisciplinary Partnership to Accelerate Community Transformation (IC-IMPACTS) Canada for their funding support. Additionally, the authors are thankful to the Federal Office of the Environment (FOEN) of Switzerland (https://www.bafu.admin.ch/bafu/en/home/topics/water/state.html) for providing us with daily river water temperature and low flow data for Switzerland.

Author contributions

Shahid L: Conceptualization, Methodology, Software, Formal analysis, Validation, Writing-original draft preparation, Project administration**. Taha B.M.J Ouarda:** Project Focus and Supervision, Funding acquisition, Conceptualization, Methodology, Project administration, Writing-Review & editing, Results validations**, André- St-Hilaire:** Project focus, Conceptualization, Writing-Review & editing, Results validations, **Zina Souaissi:** Database preparation, Writing-Review & editing, **Shaik Rehana:** Writing-Review & editing

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this manuscript.

Data Availability Statement

The datasets used in this study, for instance, daily-basis river water temperature and river low flow data for the selected river stations in Switzerland, are available at https://www.bafu.admin.ch/bafu/en/home/topics/water/state.html (Federal Office of the Environment (FOEN) of Switzerland). The R-software used in this study for data analysis and modelling is freely available at https://www.R-project.org/.

References:

Abidi, O., St-Hilaire, A., Ouarda, T. B. M. J., Charron, C., Boyer, C., & Daigle, A. (2022). Regional thermal analysis approach: A management tool for predicting water temperature metrics relevant for thermal fish habitat. Ecological Informatics, 70, 101692. https://doi.org/10.1016/j.ecoinf.2022.101692.

Adamowski, K. (1989). A Monte Carlo comparison of parametric and nonparametric estimation of flood frequencies. Journal of Hydrology, 108, 295–308. https://doi.org/10.1016/0022-1694(89)90290-4.

Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19(6), 716–723. https://doi.org/10.1109/tac.1974.1100705.

Alobaidi, M. H., Ouarda, T. B. M. J., Marpu, P. R., & Chebana, F. (2021). Diversity-driven ANN-based ensemble framework for seasonal low-flow analysis at ungauged sites. Advances in Water Resources, 147, 103814. https://doi.org/10.1016/j.advwatres.2020.103814.

Behnen, K., Hušková, M., & Neuhaus, G. (1985). Rank Estimators of scores for testing independence. Statistics & Risk Modeling, 3(3–4). https://doi.org/10.1524/strm.1985.3.34.239.

Bennett, N. D., Croke, B. F. W., Guariso, G., Guillaume, J. H. A., Hamilton, S. H., Jakeman, A. J., Marsili-Libelli, S., Newham, L. T. H., Norton, J. P., Perrin, C., Pierce, S. A., Robson, B., Seppelt, R., Voinov, A. A., Fath, B. D., & Andreassian, V. (2013). Characterizing performance of environmental models. Environmental Modelling & Software, 40, 1–20. https://doi.org/10.1016/j.envsoft.2012.09.011.

Boldi, M.-O., & Davison, A. C. (2007). A Mixture Model for Multivariate Extremes. Journal of the Royal Statistical Society Series B: Statistical Methodology, 69(2), 217–229. https://doi.org/10.1111/j.1467-9868.2007.00585.x.

Booker, D. J., & Whitehead, A. L. (2021). River water temperatures are higher during lower flows after accounting for meteorological variability. River Research and Applications, 38(1), 3–22. Portico. https://doi.org/10.1002/rra.3870.

Bouezmarni, T., & Rombouts, J. V. K. (2009). Semiparametric multivariate density estimation for positive data using copulas. Computational Statistics & Data Analysis, 53(6), 2040–2054. https://doi.org/10.1016/j.csda.2008.06.005..

Bouezmarni, T., Ghouch, E., & Taamouti, A. (2013). Bernstein estimator for unbounded copula densities. Statistics & Risk Modeling, 30(4), 343–360. https://doi.org/10.1524/strm.2013.2003.

Brown, B. M., & Chen, S. X. (1999). Beta-Bernstein Smoothing for Regression Curves with Compact Support. Scandinavian Journal of Statistics, 26(1), 47–59. https://doi.org/10.1111/1467-9469.00136.

Caissie, D., Satish, M. G., & El-Jabi, N. (2007). Predicting water temperatures using a deterministic model: Application on Miramichi River catchments (New Brunswick, Canada). Journal of Hydrology, 336(3–4), 303–315. https://doi.org/10.1016/j.jhydrol.2007.01.008.

Caissie, D., Ashkar, F., & El-Jabi, N. (2019). Analysis of air/river maximum daily temperature characteristics using the peaks over threshold approach. Ecohydrology, 13(1). Portico. https://doi.org/10.1002/eco.2176.CAISSIE, D. (2006). The thermal regime of rivers: a review. Freshwater Biology, 51(8), 1389–1406. https://doi.org/10.1111/j.1365-2427.2006.01597.x.

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. Geoscientific Model Development, 7(3), 1247–1250. https://doi.org/10.5194/gmd-7-1247-2014.

Charpentier A, Fermanian J, Scaillet O (2006) Copulas: from theory to application in finance, 1st edn, Risk Books, Torquay, UK, chap The Estimation of Copulas: Theory and Practice.

Chebana, F., & Ouarda, T. B. M. J. (2021). Multivariate nonstationary hydrological frequency analysis. Journal of Hydrology, 593, 125907. https://doi.org/10.1016/j.jhydrol.2020.125907.

Chen, S. X. (1999). Beta kernel estimators for density functions. Computational Statistics & Data Analysis, 31(2), 131–145. https://doi.org/10.1016/s0167-9473(99)00010-9.

Chen, S. X., & Huang, T.-M. (2007). Nonparametric estimation of copula functions for dependence modelling. Canadian Journal of Statistics, 35(2), 265–282. https://doi.org/10.1002/cjs.5550350205.

Chen, S. (2015). Optimal Bandwidth Selection for Kernel Density Functionals Estimation. Journal of Probability and Statistics, 2015, 1–21. https://doi.org/10.1155/2015/242683.

Coles, S. G., & Tawn, J. A. (1994). Statistical Methods for Multivariate Extremes: An Application to Structural Design. Applied Statistics, 43(1), 1. https://doi.org/10.2307/2986112.

Coles, S., J. Heffernan, and J. Tawn (1999), Dependence measures for extreme value analyses, Extremes, 2(4), 339–365.

Coles, S. (2001). An Introduction to Statistical Modeling of Extreme Values. Springer Series in Statistics. https://doi.org/10.1007/978-1-4471-3675-0.

Daigle, A., St-Hilaire, A., Beveridge, D., Caissie, D., & Benyahya, L. (2011). Multivariate analysis of the low-flow regimes in eastern Canadian rivers. Hydrological Sciences Journal, 56(1), 51–67. https://doi.org/10.1080/02626667.2010.535002.

Deheuvels, P. and P. Hominal, (1979). Estimation non paramétrique de la densité compte tenu d'informations sur le support, Revue de Statistique Appliquée, 27, pp 47–68.

Devroye, L. and Györfi, L (1985). Nonparametric Density Estimation: The L 1 View, Wiley, New York.

Diers, D., Eling, M., & Marek, S. D. (2012). Dependence modeling in non-life insurance using the Bernstein copula. Insurance: Mathematics and Economics, 50(3), 430–436. https://doi.org/10.1016/j.insmatheco.2012.02.007.

Elliott, & Hurley. (2001). Modelling growth of brown trout, Salmo trutta, in terms of weight and energy units. Freshwater Biology, 46(5), 679–692. Portico. https://doi.org/10.1046/j.1365-2427.2001.00705.x.

Farrell, P. J., & Rogers-Stewart, K. (2006). Comprehensive study of tests for normality and symmetry: extending the Spiegelhalter test. Journal of Statistical Computation and Simulation, 76(9), 803–816. https://doi.org/10.1080/10629360500109023.

Ficklin, D. L., Stewart, I. T., & Maurer, E. P. (2013). Effects of climate change on stream temperature, dissolved oxygen, and sediment concentration in the Sierra Nevada in California. Water Resources Research, 49(5), 2765–2782. Portico. https://doi.org/10.1002/wrcr.20248.

Fullerton, A. H., Burnett, K. M., Steel, E. A., Flitcroft, R. L., Pess, G. R., Feist, B. E., Torgersen, C. E., Miller D. J., & Sanderson, B. L. (2010). Hydrological connectivity for riverine fish: measurement challenges and research opportunities. Freshwater Biology, 55(11), 2215–2237. Portico. https://doi.org/10.1111/j.1365-2427.2010.02448.x.

Geenens, G., Charpentier, A., & Paindaveine, D. (2017). Probit transformation for nonparametric kernel estimation of the copula density. Bernoulli, 23(3). https://doi.org/10.3150/15-bej798.

Genest, C. K. G. and L. Rivest (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. Biometrika, 82(3), 543–552. https://doi.org/10.1093/biomet/82.3.543.

Gijbels, I., & Mielniczuk, J. (1990). Estimating the density of a copula function. Communications in Statistics - Theory and Methods, 19(2), 445–464. https://doi.org/10.1080/03610929008830212.

Goel NK, Seth SM, Chandra S (1998) Multivariate Modelling of flood flows. J Hydraul Eng 124(2):146–155. https://doi.org/10.1061/(ASCE)0733-9429(1998)124:2(146).

Gräler, B., van den Berg, M. J., Vandenberghe, S., Petroselli, A., Grimaldi, S., De Baets, B., & Verhoest, N. E. C. (2013). Multivariate return periods in hydrology: a critical and practical review focusing on synthetic design hydrograph estimation. Hydrology and Earth System Sciences, 17(4), 1281–1296. https://doi.org/10.5194/hess-17-1281-2013.

Gringorten, I. I. (1963). A plotting rule for extreme probability paper. Journal of Geophysical Research, 68(3), 813–814. https://doi.org/10.1029/jz068i003p00813.

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological Modelling. Journal of Hydrology, 377(1–2), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003.

Haggag, M. M. M. (2014). New Criteria of Model Selection and Model Averaging in Linear Regression Models. American Journal of Theoretical and Applied Statistics, 3(5), 148. https://doi.org/10.11648/j.ajtas.20140305.15

Hannan, E. J., & Quinn, B. G. (1979). The Determination of the Order of an Autoregression. Journal of the Royal Statistical Society: Series B (Methodological), 41(2), 190–195. Portico. https://doi.org/10.1111/j.2517-6161.1979.tb01072.x.

Han, Q., & Chu, F. (2021). Directional wind energy assessment of China based on nonparametric copula models. Renewable Energy, 164, 1334–1349. https://doi.org/10.1016/j.renene.2020.10.149.

Härdle, W. (1991). Kernel Density Estimation. In: Smoothing Techniques. Springer Series in Statistics. Springer, New York, NY. https://doi.org/10.1007/978-1-4612-4432-5_2.

HARRELL, F. E., & DAVIS, C. E. (1982). A new distribution-free quantile estimator. Biometrika, 69(3), 635–640. https://doi.org/10.1093/biomet/69.3.635.

Heffernan, J. E., & Tawn, J. A. (2004). A Conditional Approach for Multivariate Extreme Values (with Discussion). Journal of the Royal Statistical Society Series B: Statistical Methodology, 66(3), 497–546. https://doi.org/10.1111/j.1467-9868.2004.02050.x.

Hendry, A., Haigh, I. D., Nicholls, R. J., Winter, H., Neal, R., Wahl, T., Joly-Laugel, A., and Darby, S. E. (2019) Assessing the characteristics and drivers of compound flooding events around the UK coast, Hydrol. Earth Syst. Sci., 23, 3117–3139, https://doi.org/10.5194/hess-23-3117-2019.

Humphries, P., & Baldwin, D. S. (2003). Drought and aquatic ecosystems: an introduction. Freshwater Biology, 48(7), 1141–1146. Portico. https://doi.org/10.1046/j.1365-2427.2003.01092.x.

Joe, H. (1997). Multivariate Models and Multivariate Dependence Concepts. C&H/CRC Monographs on Statistics & Applied Probability. https://doi.org/10.1201/b13150.

Jones, M. C., Marron, J. S., & Sheather, S. J. (1996). A Brief Survey of Bandwidth Selection for Density Estimation. Journal of the American Statistical Association, 91(433), 401–407. https://doi.org/10.1080/01621459.1996.10476701.

Joshi, D., St-Hilaire, A., Ouarda, T. B. M. J., Daigle, A., & Thiemonge, N. (2016). Comparison of direct statistical and indirect statistical-deterministic frameworks in downscaling river low-flow indices. Hydrological Sciences Journal, 61(11), 1996–2010. https://doi.org/10.1080/02626667.2014.966719.

Karmakar S, Simonovic SP (2009) Bivariate flood frequency analysis. Part-2: a copula-based approach with mixed marginal distributions. J Flood Risk Manag 2(1):1–13. https://doi.org/10.1111/j.1753-318X.2009.01020.x.

Kim, T.-W., Valdés, J. B., & Yoo, C. (2006). Nonparametric Approach for Bivariate Drought Characterization Using Palmer Drought Index. Journal of Hydrologic Engineering, 11(2), 134–143. https://doi.org/10.1061/(asce)1084-0699(2006)11:2(134).

Körner, O., Kohno, S., Schönenberger, R., Suter, M. J.-F., Knauer, K., Guillette, L. J., & Burkhardt-Holm, P. (2008). Water temperature and concomitant waterborne ethinylestradiol exposure affects the vitellogenin expression in juvenile brown trout (Salmo trutta). Aquatic Toxicology, 90(3), 188–196. https://doi.org/10.1016/j.aquatox.2008.08.012.

Krause, P., Boyle, D. P., & Bäse, F. (2005). Comparison of different efficiency criteria for hydrological model assessment. Advances in Geosciences, 5, 89–97. https://doi.org/10.5194/adgeo-5-89-2005.

Kulpa, T. (1999). On approximation of copulas. International Journal of Mathematics and Mathematical Sciences, 22(2), 259–269. https://doi.org/10.1155/s0161171299222594.

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. Journal of Hydrology, 10(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6.

Latif, S., Souaissi, Z., & Ouarda, T. B. (2023). Copula-based joint Modelling of extreme river temperature and low flow characteristics in the risk assessment of aquatic life. Weather and Climate Extremes, 100586. https://doi.org/10.1016/j.wace.2023.100586.

Latif, S., & Mustafa, F. (2020a). Trivariate distribution modelling of flood characteristics using copula function—A case study for Kelantan River basin in Malaysia. AIMS Geosciences, 6(1), 92–130. https://doi.org/10.3934/geosci.2020007.

Latif, S., & Mustafa, F. (2020b). A nonparametric copula distribution framework for bivariate joint distribution analysis of flood characteristics for the Kelantan River basin in Malaysia. AIMS Geosciences, 6(2), 171–198. https://doi.org/10.3934/geosci.2020012.

Latif, S., & Mustafa, F. (2021). Bivariate joint distribution analysis of the flood characteristics under semiparametric copula distribution framework for the Kelantan River basin in Malaysia. Journal of Ocean Engineering and Science, 6(2), 128–145. https://doi.org/10.1016/j.joes.2020.06.003.

Latif, S., & Simonovic, S. P. (2022a). Trivariate Joint Distribution Modelling of Compound Events Using the Nonparametric D-Vine Copula Developed Based on a Bernstein and Beta Kernel Copula Density Framework. Hydrology, 9(12), 221. https://doi.org/10.3390/hydrology9120221.

Latif, S., & Simonovic, S. P. (2022b). Parametric Vine Copula Framework in the Trivariate Probability Analysis of Compound Flooding Events. Water, 14(14), 2214. https://doi.org/10.3390/w14142214.

Latif, S., & Simonovic, S. P. (2022c). Nonparametric Approach to Copula Estimation in Compounding The Joint Impact of Storm Surge and Rainfall Events in Coastal Flood Analysis. Water Resources Management, 36(14), 5599–5632. https://doi.org/10.1007/s11269-022-03321-y.

Lee, T., Ouarda, T. B. M. J., & Yoon, S. (2017). KNN-based local linear regression for the analysis and simulation of low flow extremes under climatic influence. Climate Dynamics, 49(9–10), 3493–3511. https://doi.org/10.1007/s00382-017-3525-0.

Legates, D. R., & McCabe, G. J. (1999). Evaluating the use of "goodness-of-fit" Measures in hydrologic and hydroclimatic model validation. Water Resources Research, 35(1), 233–241. Portico. https://doi.org/10.1029/1998wr900018.

Liang, Y., Wu, C., Zhang, M., Ji, X., Shen, Y., He, J., & Zhang, Z. (2022). Statistical Modelling of the joint probability density function of air density and wind speed for wind resource assessment: A case study from China. Energy Conversion and Management, 268, 116054. https://doi.org/10.1016/j.enconman.2022.116054.

Lund, S. G., Caissie, D., Cunjak, R. A., Vijayan, M. M., & Tufts, B. L. (2002). The effects of environmental heat stress on heat-shock mRNA and protein expression in Miramichi Atlantic salmon (Salmo salar) parr. Canadian Journal of Fisheries and Aquatic Sciences, 59(9), 1553–1562. https://doi.org/10.1139/f02-117.

Michel, A., Brauchli, T., Lehning, M., Schaefli, B., & Huwald, H. (2020). Stream temperature and discharge evolution in Switzerland over the last 50 years: annual and seasonal behaviour. Hydrology and Earth System Sciences, 24(1), 115–142. https://doi.org/10.5194/hess-24-115-2020.

Moftakhari, H. R., Salvadori, G., AghaKouchak, A., Sanders, B. F., & Matthew, R. A. (2017). Compounding effects of sea level rise and fluvial flooding. Proceedings of the National Academy of Sciences, 114(37), 9785–9790. https://doi.org/10.1073/pnas.1620325114.

Molanes-López, E. M., & Cao, R. (2007). Plug-in bandwidth selector for the kernel relative density estimator. Annals of the Institute of Statistical Mathematics, 60(2), 273–300. https://doi.org/10.1007/s10463-006-0108-y.

Moriasi, D.N., J. G. Arnold, M. W. Van Liew, R. L. Bingner, R. D. Harmel, & T. L. Veith. (2007). Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. Transactions of the ASABE, 50(3), 885–900. https://doi.org/10.13031/2013.23153.

MÜLLER, H.-G. (1991). Smooth optimum kernel estimators near endpoints. Biometrika, 78(3), 521–530. https://doi.org/10.1093/biomet/78.3.521.

Nagler, T. (2014). Kernel Methods for Vine Copula Estimation. Master's Thesis, Technische Universitaet Muenchen, https://mediatum.ub.tum.de/node?id=1231221.

Nelsen RB (2006) An introduction to copulas. Springer, New York.

Onyutha, C. (2021). A hydrological model skill score and revised R-squared. Hydrology Research, 53(1), 51–64. https://doi.org/10.2166/nh.2021.071.

Ouarda, T. B. M. J., Charron, C., Shin, J.-Y., Marpu, P. R., Al-Mandoos, A. H., Al-Tamimi, M. H., Ghedira, H., & Al Hosary, T. N. (2015). Probability distributions of wind speed in the UAE. Energy Conversion and Management, 93, 414–434. https://doi.org/10.1016/j.enconman.2015.01.036.

Ouarda, T. B. M. J., Charron, C., Hundecha, Y., St-Hilaire, A., & Chebana, F. (2018). Introduction of the GAM model for regional low-flow frequency analysis at ungauged basins and comparison with commonly used approaches. Environmental Modelling & Software, 109, 256–271. https://doi.org/10.1016/j.envsoft.2018.08.031.

Ouarda, T. B. M. J., Charron, C., & St-Hilaire, A. (2022). Regional estimation of river water temperature at ungauged locations. Journal of Hydrology X, 17, 100133. https://doi.org/10.1016/j.hydroa.2022.100133.

Pfeifer, D., Strassburger, D. and Philipps, J (2009): "Modelling and simulation of dependence structures in nonlife insurance with Bernstein copulas," Working Paper, Carl von Ossietzky University, Oldenburg.

Rauf, U.F.A., Zeephongsekul, P. Analysis of Rainfall Severity and Duration in Victoria, Australia using Nonparametric Copulas and Marginal Distributions. Water Resour Manage 28, 4835–4856 (2014). https://doi.org/10.1007/s11269-014-0779-8.

Reddy MJ, Ganguli P (2012) Bivariate flood frequency analysis of Upper Godavari River flows using Archimedean copulas. Water Resour Manage: DOI. https://doi.org/10.1007/s11269-012-0124-z.

Renault, O., & Scaillet, O. (2004). On the way to recovery: A nonparametric bias free estimation of recovery rate densities. Journal of Banking & Finance, 28(12), 2915–2931. https://doi.org/10.1016/j.jbankfin.2003.10.018.

Rosenblatt, M. (1956). Remarks on Some Nonparametric Estimates of a Density Function. The Annals of Mathematical Statistics, 27(3), 832–837. https://doi.org/10.1214/aoms/1177728190.

Saklar A (1959) Functions de repartition n dimensions et leurs marges. Publ Inst Stat Univ Paris 8:229–231.

Salvadori G (2004). Bivariate return periods via-2 copulas. J Royal Stat Soc Series B 1:129–144. https://doi.org/10.1016/j.stamet.2004.07.002.

Salvadori G, De Michele C (2010) Multivariate multiparameters extreme value models and return periods: a Copula approach. Water Resour Res. https://doi.org/10.1029/2009WR009040.

Salvadori G, De Michele C, Durante F (2011) Multivariate design via copulas. Hydrol Earth Sys Sci Discuss 8(3):5523–5558. https://doi.org/10.5194/hessd-8-5523-2011.

Salvadori G, Durante F, Tomasicchio GR, D'Alessandro F (2015) Practical guidelines for the multivariate assessments of the structural risk in coastal and offshore engineering. Coast Engg 95:77–83. https://doi.org/10.1016/j.coastaleng.2014.09.007.

Samuels, P. G., & Burt, N. (2002). A new joint probability appraisal of flood risk. Proceedings of the Institution of Civil Engineers - Water and Maritime Engineering, 154(2), 109–115. https://doi.org/10.1680/wame.2002.154.2.109.

Sancetta, A., & Satchell, S. (2004). THE BERNSTEIN COPULA AND ITS APPLICATIONS TO MODELING AND APPROXIMATIONS OF MULTIVARIATE DISTRIBUTIONS. Econometric Theory, 20(03). https://doi.org/10.1017/s026646660420305x.

Santhosh D, Srinivas VV (2013) Bivariate frequency analysis of flood using a diffusion kernel density estimators. Water Resour Res 49:8328–8343. https://doi.org/10.1002/2011WR0100777.

Schuster, E. F. (1985). Incorporating support constraints into nonparametric estimators of densities. Communications in Statistics - Theory and Methods, 14(5), 1123–1136. https://doi.org/10.1080/03610928508828965.

Schwarz, G. (1978). Estimating the Dimension of a Model. The Annals of Statistics, 6(2). https://doi.org/10.1214/aos/1176344136.

Scott, D. W. (1992). Multivariate Density Estimation. Wiley Series in Probability and Statistics. https://doi.org/10.1002/9780470316849.

Seneviratne, S., Nicholls, N., Easterling, D., Goodess, C., Kanae, S., Kossin, J., Luo, Y., Marengo, J., McInnes, K., Rahimi, M., Reichstein, M., Sorteberg, A., Vera, C., and Zhang, X. (2012) Changes in climate extremes and their impacts on the natural physical environment, Manag. Risk Extrem. Events Disasters to Adv. Clim. Chang. Adapt., 109–230, available at: https://www. ipcc.ch/pdf/special-reports/srex/SREXChap3_ FINAL.pdf.

Sharma, A., Lall, U., & Tarboton, D. G. (1998). Kernel bandwidth selection for a first order nonparametric streamflow simulation model. Stochastic Hydrology and Hydraulics, 12(1), 33–52. https://doi.org/10.1007/s004770050008.

Sheather, S. J., & Jones, M. C. (1991). A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation. Journal of the Royal Statistical Society: Series B (Methodological), 53(3), 683–690. Portico. https://doi.org/10.1111/j.2517-6161.1991.tb01857.x.

Shiau, J. T. (2006). Fitting Drought Duration and Severity with Two-Dimensional Copulas. Water Resources Management, 20(5), 795–815. https://doi.org/10.1007/s11269-005-9008-9.

Shih, J. H., & Louis, T. A. (1995). Inferences on the Association Parameter in Copula Models for Bivariate Survival Data. Biometrics, 51(4), 1384. https://doi.org/10.2307/2533269.

Silverman B. W. (1986). Density Estimation for Statistics and Data Analysis, 1st edn. Chapman and Hall, London.

Singh, J., Knapp, H.V. and Demissie, M. (2004), "Hydrologic modeling of the Iroquois River watershed using HSPF and SWAT. ISWS CR 2004-08. Champaign, Ill.: Illinois State Water Survey. Available at: www.sws.uiuc.edu/pubdoc/CR/ ISWSCR2004-08.pdf. Accessed September 8 2005.

Sinokrot, B. A., & Gulliver, J. S. (2000). In-stream flow impact on river water temperatures. Journal of Hydraulic Research, 38(5), 339–349. https://doi.org/10.1080/00221680009498315.

Sorooshian, S., Duan, Q., & Gupta, V. K. (1993). Calibration of rainfall-runoff models: Application of global optimization to the Sacramento Soil Moisture Accounting Model. Water Resources Research, 29(4), 1185–1194. Portico. https://doi.org/10.1029/92wr02617.

Souaissi, Z., Ouarda, T. B. M. J., & St-Hilaire, A. (2021). River water temperature quantiles as thermal stress indicators: Case study in Switzerland. Ecological Indicators, 131, 108234. https://doi.org/10.1016/j.ecolind.2021.108234.

St-Hilaire, A., Ouarda, T. B. M. J., Bargaoui, Z., Daigle, A., & Bilodeau, L. (2011). Daily river water temperature forecast model with a k-nearest neighbour approach. Hydrological Processes, 26(9), 1302–1310. https://doi.org/10.1002/hyp.8216.

Strepparava, N., Segner, H., Ros, A., Hartikainen, H., Schmidt-Posthaus, H., & Wahli, T. (2017). Temperature-related parasite infection dynamics: the case of proliferative kidney disease of brown trout. Parasitology, 145(3), 281–291. https://doi.org/10.1017/s0031182017001482.

Svensson, C. and Jones, D. A. (2004) Dependence between sea surge, river flow and precipitation in south and west Britain, Hydrol. Earth Syst. Sci., 8, 973–992, https://doi.org/10.5194/hess-8-973-2004.

Vernieuwe H, Vandenberghe S, Baets BD, Verhost NEC (2015) A continuous rainfall model based on vine copulas. Hydrol Earth Syst Sci 19:2685–2699. https://doi.org/10.5194/hess-19-2685-2015.

Vitale, R. A. (1975). A Bernstein Polynomial Approach to Density Function Estimation. Statistical Inference and Related Topics, 87–99. https://doi.org/10.1016/b978-0-12-568002-8.50011-2.

Wahl, T., Jain, S., Bender, J., Meyers, S. D., & Luther, M. E. (2015). Increasing risk of compound flooding from storm surge and rainfall for major US cities. Nature Climate Change, 5(12), 1093–1097. https://doi.org/10.1038/nclimate2736.

Wand, M. P., Marron, J. S., & Ruppert, D. (1991). Transformations in Density Estimation: Rejoinder. Journal of the American Statistical Association, 86(414), 360. https://doi.org/10.2307/2290575.

Wand, M. P., & Jones, M. C. (1995). Kernel Smoothing. https://doi.org/10.1007/978-1-4899-4493-1.

Weiss, G. N. F., & Scheffer, M. (2012). Smooth Nonparametric Bernstein Vine Copulas. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.2154458.

Willmott, C. J. (1984). On the Evaluation of Model Performance in Physical Geography. Spatial Statistics and Models, 443–460. https://doi.org/10.1007/978-94-017-3048-8_23.

Willmott, C. J., Ackleson, S. G., Davis, R. E., Feddema, J. J., Klink, K. M., Legates, D. R., O'Donnell, J., & Rowe, C. M. (1985). Statistics for the evaluation and comparison of models. Journal of Geophysical Research, 90(C5), 8995. https://doi.org/10.1029/jc090ic05p08995.

Willmott, C., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Research, 30, 79–82. https://doi.org/10.3354/cr030079.

Yue, S., Ouarda, T. B. M. J., Bobée, B., Legendre, P., & Bruneau, P. (1999). The Gumbel mixed model for flood frequency analysis. Journal of Hydrology, 226(1–2), 88–100. https://doi.org/10.1016/s0022-1694(99)00168-7.

Zhang, L., & Singh, V. P. (2007). Trivariate Flood Frequency Analysis Using the Gumbel–Hougaard Copula. Journal of Hydrologic Engineering, 12(4), 431–439. https://doi.org/10.1061/(asce)1084-0699(2007)12:4(431).

Zscheischler, J., Westra, S., van den Hurk, B. J. J. M., Seneviratne, S. I., Ward, P. J., Pitman, A., AghaKouchak, A., Bresch, D. N., Leonard, M., Wahl, T., & Zhang, X. (2018). Future climate risk from compound events. Nature Climate Change, 8(6), 469–477. https://doi.org/10.1038/s41558-018-0156-3.

List of Figures

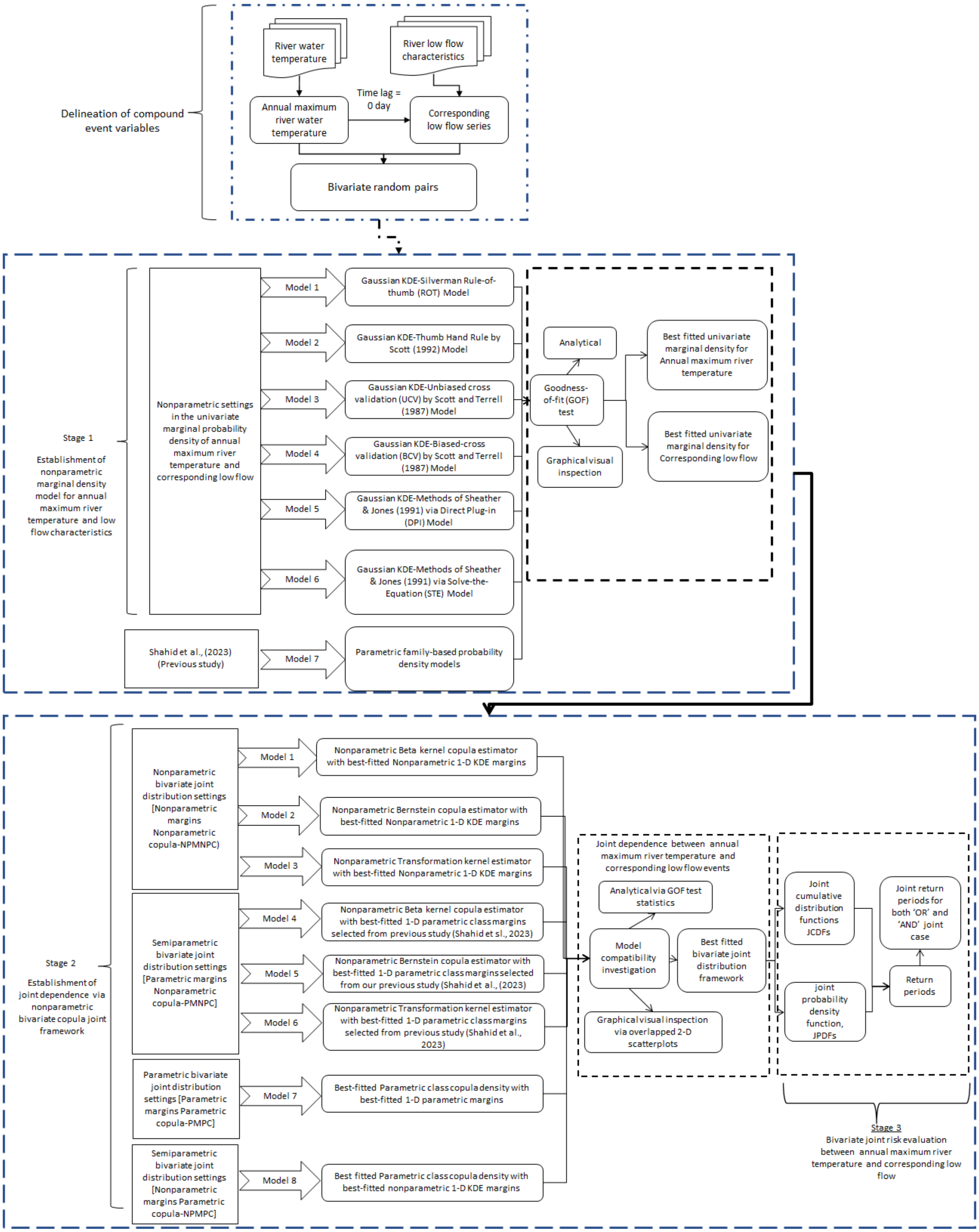


Figure 1: Methodological workflow in the nonparametric copula-based joint density modelling for annual maximum river water temperature and corresponding low flow events

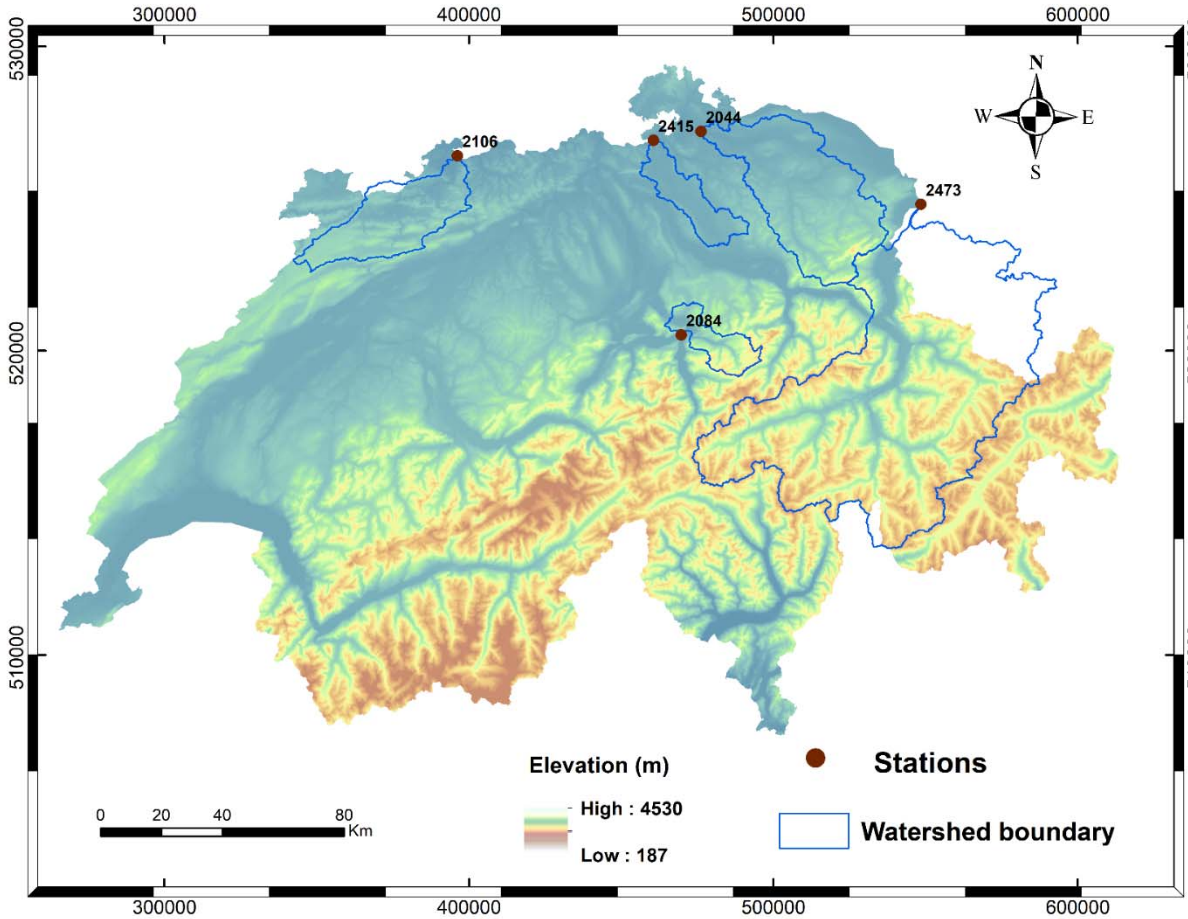
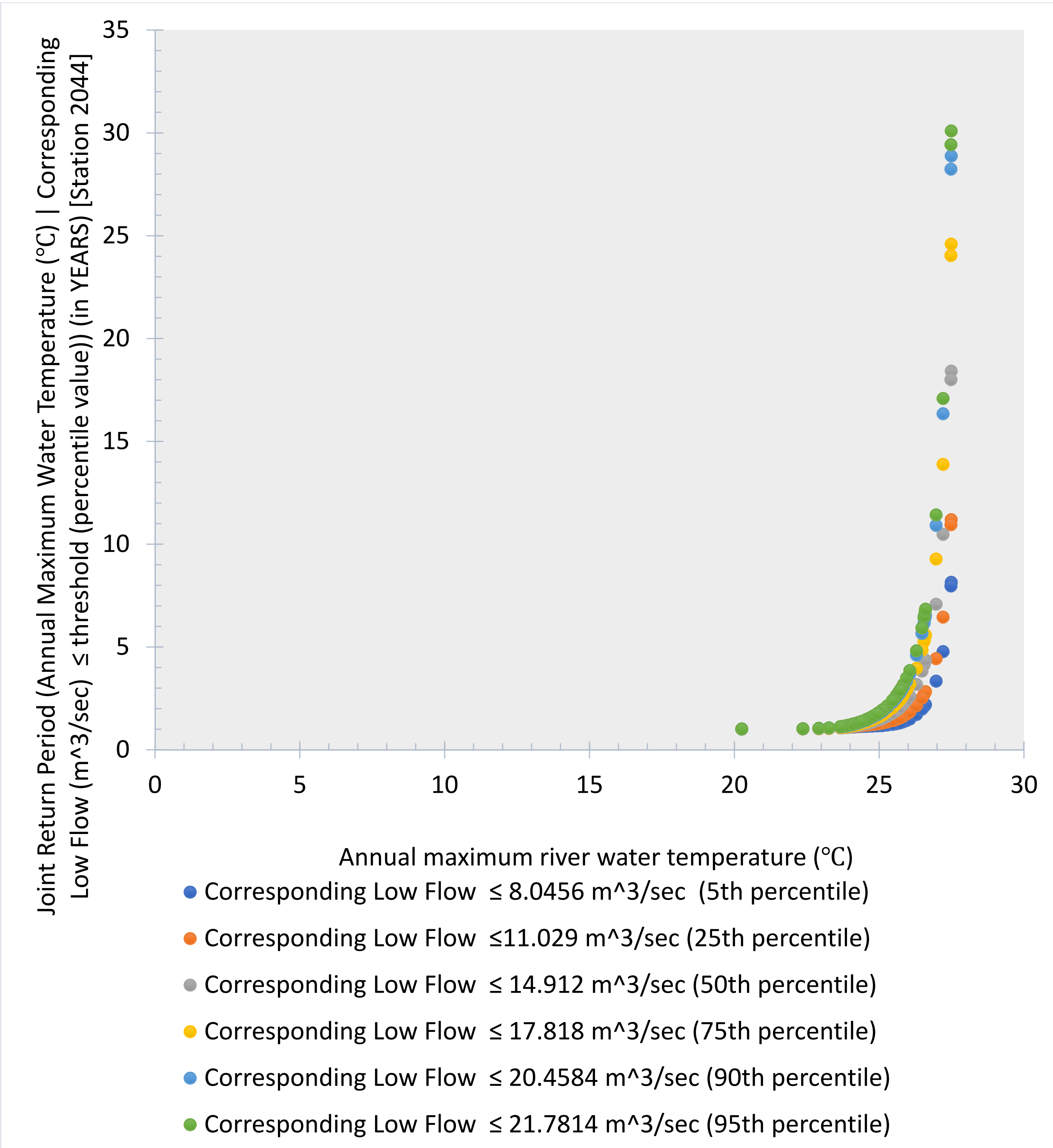
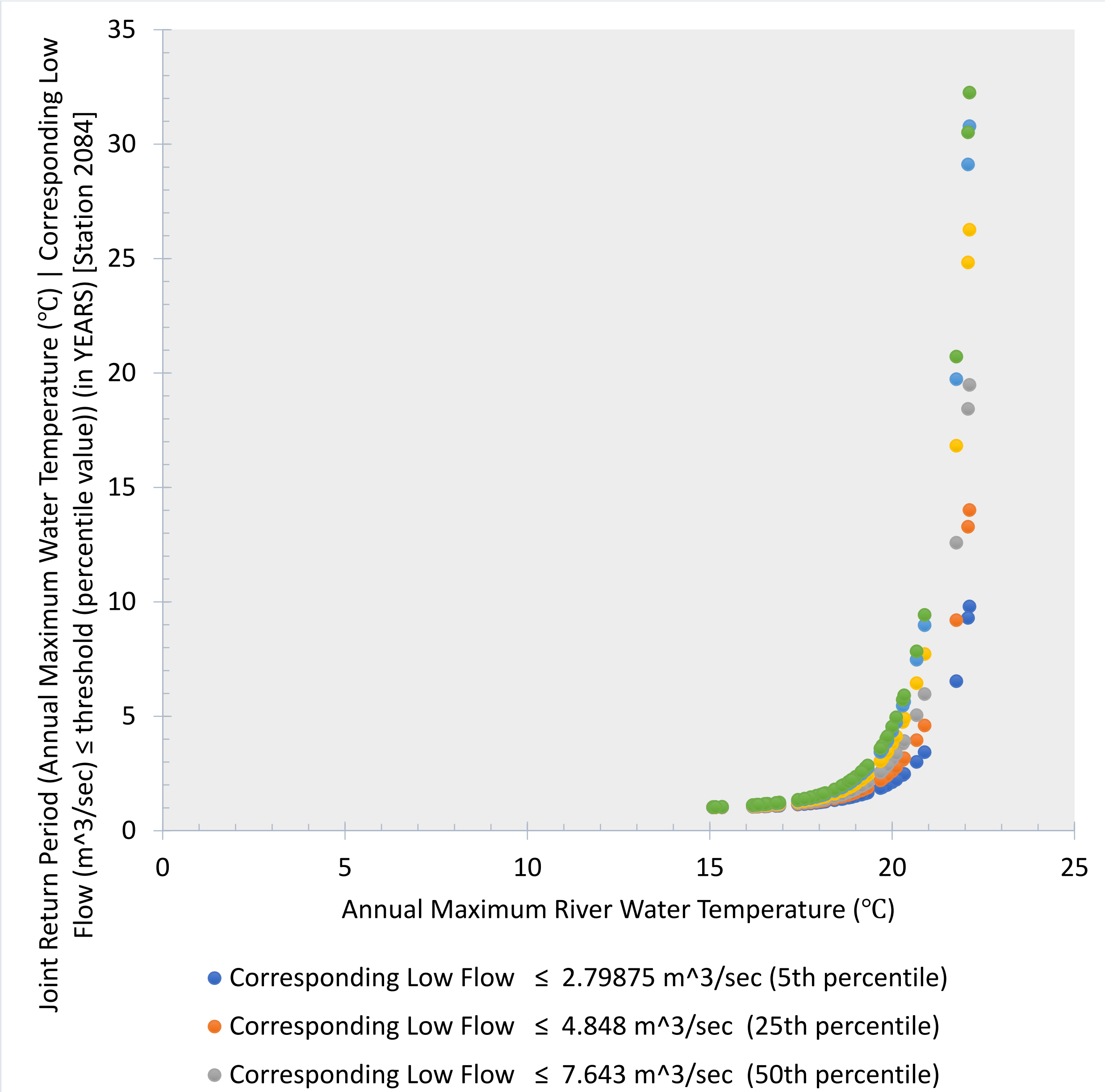


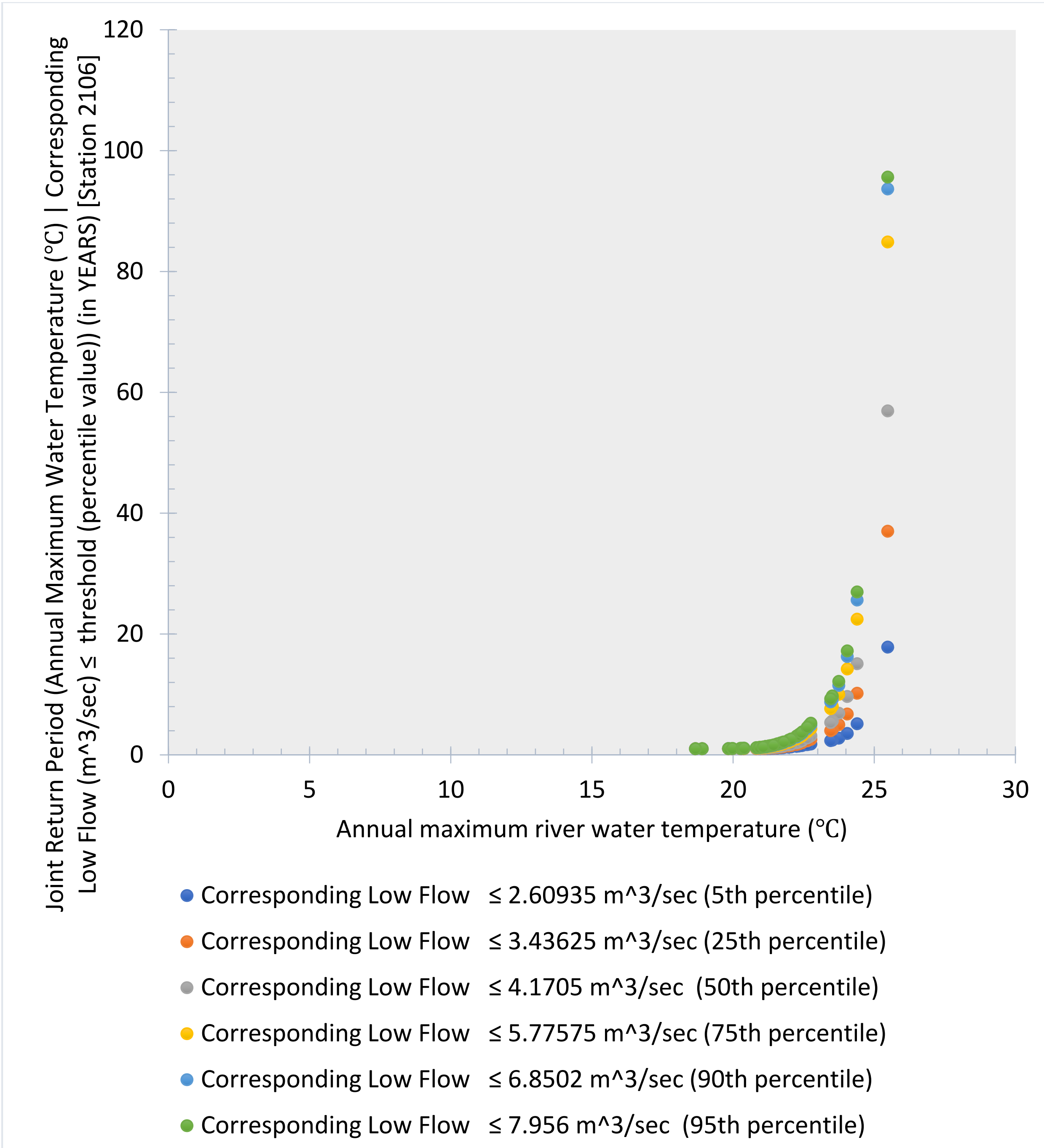
Figure 2: Geographical location of study area with river water gauge stations


Joint Return Period (Annual Maximum Water Temperature (°C) | Corresponding Low Flow (m^3/sec) ≤ threshold (percentile value)) (in YEARS) [Station 2044]
0
5
10
15
20
25
30
35
0
5
10
15
20
25
30
Annual maximum river water temperature (°C)
Corresponding Low Flow ≤ 8.0456 m^3/sec (5th percentile)
Corresponding Low Flow ≤11.029 m^3/sec (25th percentile)
Corresponding Low Flow ≤ 14.912 m^3/sec (50th percentile)
Corresponding Low Flow ≤ 17.818 m^3/sec (75th percentile)
Corresponding Low Flow ≤ 20.4584 m^3/sec (90th percentile)
Corresponding Low Flow ≤ 21.7814 m^3/sec (95th percentile)

(a)

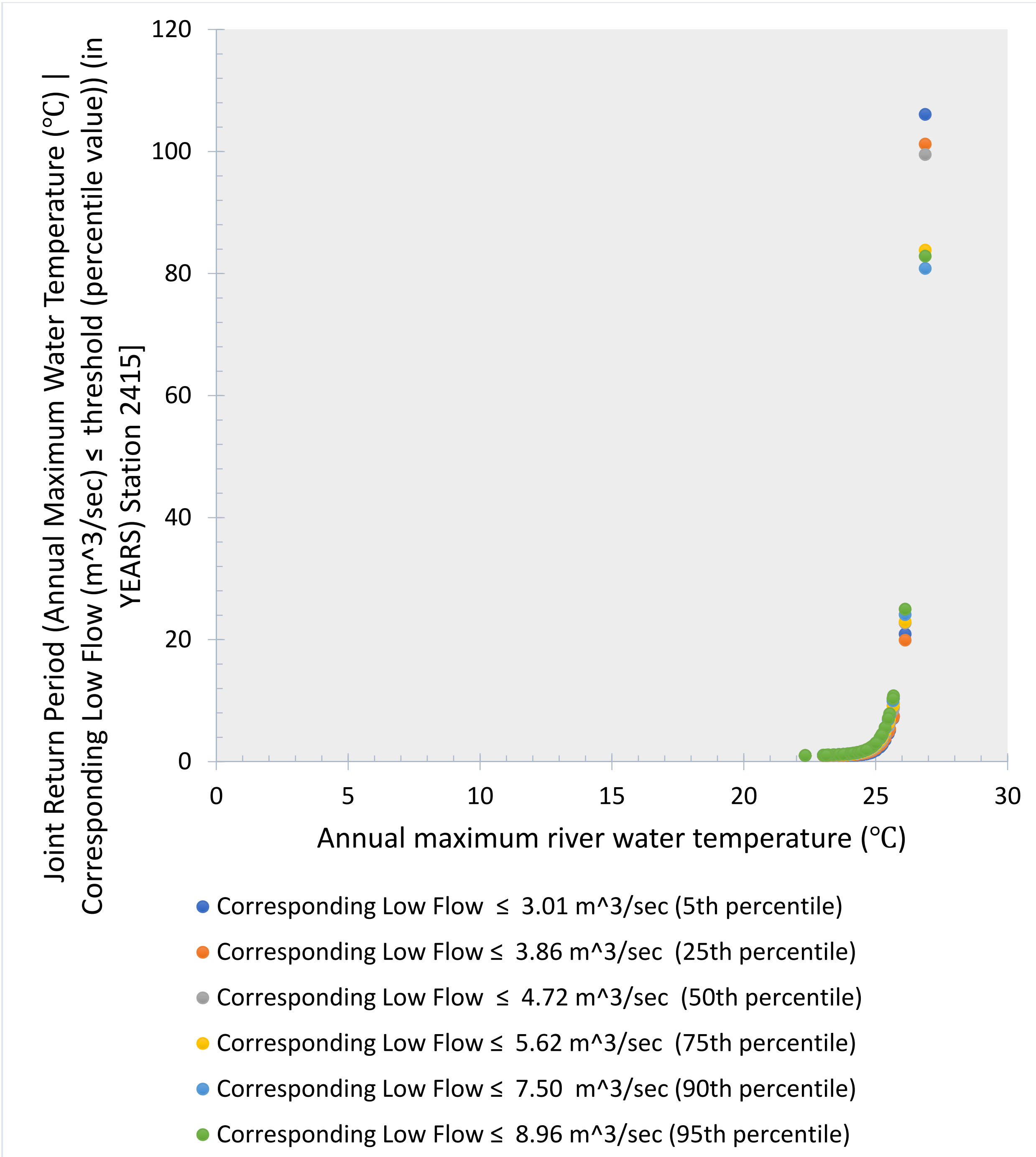

Joint Return Period (Annual Maximum Water Temperature (℃) | Corresponding Low Flow (m^3/sec) ≤ threshold (percentile value)) (in YEARS) [Station 2084]
35
30
25
20
15
10
5
0
0
5
10
15
20
25
Annual Maximum River Water Temperature (℃)
Corresponding Low Flow ≤ 2.79875 m^3/sec (5th percentile)
Corresponding Low Flow ≤ 4.848 m^3/sec (25th percentile)
Corresponding Low Flow ≤ 7.643 m^3/sec (50th percentile)
Corresponding Low Flow ≤ 13.5428 m^3/sec (75th percentile)
Corresponding Low Flow ≤ 18.263 m^3/sec (90th percentile)
Corresponding Low Flow ≤ 23.9978 m^3/sec (95th percentile)

(b)

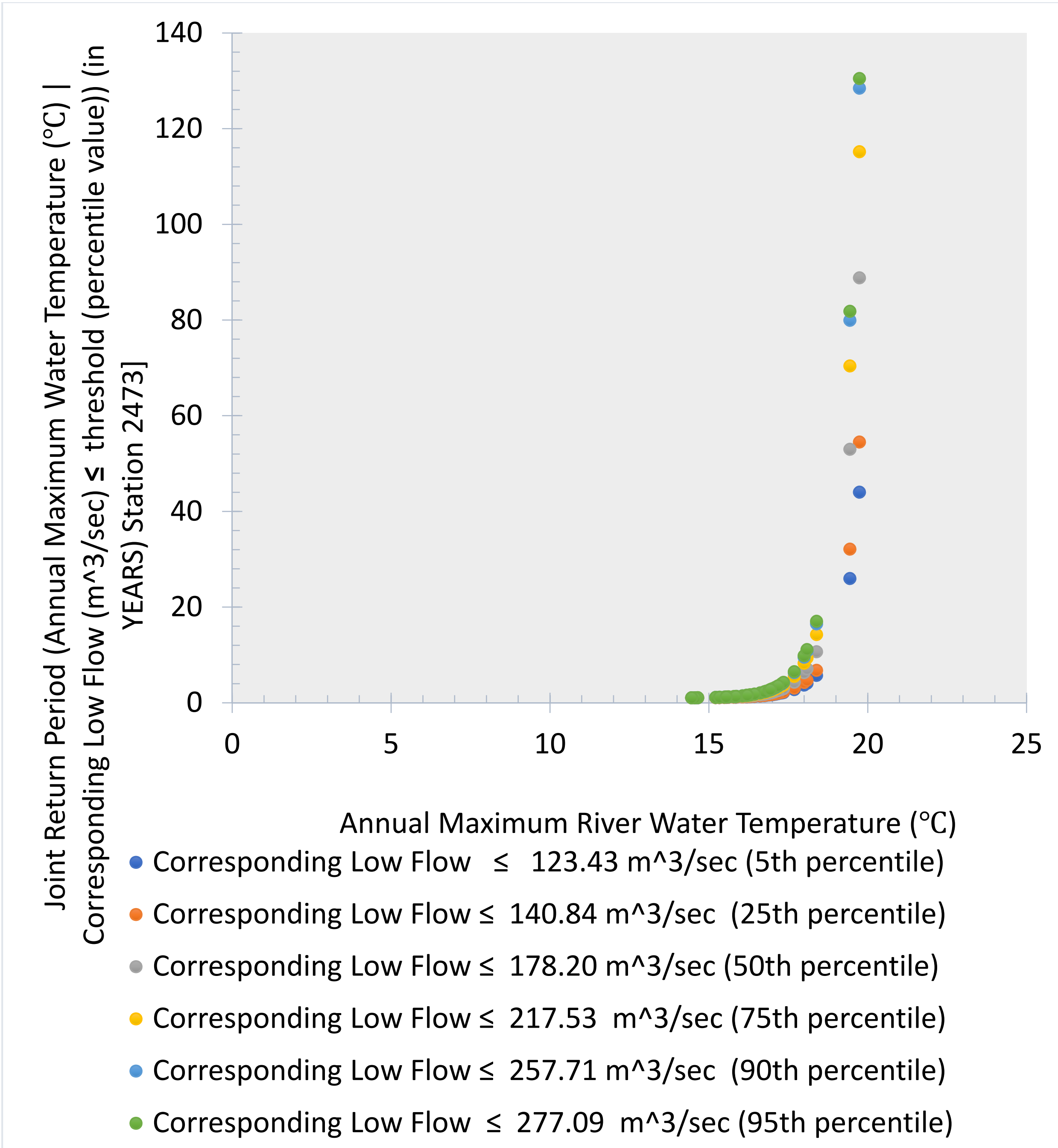


Joint Return Period (Annual Maximum Water Temperature (°C) | Corresponding Low Flow (m^3/sec) ≤ threshold (percentile value)) (in YEARS) [Station 2106]
120
100
80
60
40
20
0
0
5
10
15
20
25
30
Annual maximum river water temperature (°C)
Corresponding Low Flow ≤ 2.60935 m^3/sec (5th percentile)
Corresponding Low Flow ≤ 3.43625 m^3/sec (25th percentile)
Corresponding Low Flow ≤ 4.1705 m^3/sec (50th percentile)
Corresponding Low Flow ≤ 5.77575 m^3/sec (75th percentile)
Corresponding Low Flow ≤ 6.8502 m^3/sec (90th percentile)
Corresponding Low Flow ≤ 7.956 m^3/sec (95th percentile)

(c)


Joint Return Period (Annual Maximum Water Temperature (°C) | Corresponding Low Flow (m^3/sec) ≤ threshold (percentile value)) (in YEARS) Station 2415]
120
100
80
60
40
20
0
0
5
10
15
20
25
30
Annual maximum river water temperature (°C)
Corresponding Low Flow ≤ 3.01 m^3/sec (5th percentile)
Corresponding Low Flow ≤ 3.86 m^3/sec (25th percentile)
Corresponding Low Flow ≤ 4.72 m^3/sec (50th percentile)
Corresponding Low Flow ≤ 5.62 m^3/sec (75th percentile)
Corresponding Low Flow ≤ 7.50 m^3/sec (90th percentile)
Corresponding Low Flow ≤ 8.96 m^3/sec (95th percentile)

(d)



(e)

Figure 3. Bivariate joint return periods of annual maximum river water temperature given various percentile values of corresponding low flow for (a) station 2044 (b) station 2084 (c) station 2106 (d) station 2415 (e) station 2473

Table 1: Model compatibility or fitness investigation in fitting bivariate joint probability framework for annual maximum river water temperature and corresponding low flow series (a) station 2044 (b) station 2084 (c) station 2106 (d) station 2415 (e) station 2473

Joint distribution framework (Station 2044) (a)	Bivariate joint density frameworks	Estimated bandwidth	K-S (Kolmogorov-Smirnov)		RMSE (Root Mean Squared Error)	MSE (Mean Squared Error)	MAE (Mean Absolute Error)	NSE (Nash-Sutcliffe Efficiency)	mNSE (modified Nash-Sutcliffe Efficiency)	AIC (Akaike Information Criterion)	BIC (Bayesian Information Criterion)	HQC (Hannan-Quinn Information Criterion)	IA (Index of Agreement)	R2 (Coefficient of Determinations)	PBIAS (Percent Bias)
			Statistics	p-value											
Nonparametric margins nonparametric copula (NPMNPC) mode [Nonparametric copulas distribution settings]	**Beta kernel copula density with KDE (GAUSSIAN-Silverman Rule-of-thumb ROT) margins model***	**0.073**	**0.120**	**0.58**	**0.0200**	**0.00040**	**0.0171**	**0.980**	**0.839**	**-412.34**	**-410.37**	**-411.58**	**0.994**	**0.98**	**3.1**
	Bernstein copula with KDE (GAUSSIAN-Silverman Rule-of-thumb (ROT)) margins model	NA	0.169	0.42	0.0203	0.00041	0.0175	0.977	0.835	-410.68	-408.71	-409.92	0.993	0.98	3.2
	Transformation estimator with KDE (GAUSSIAN-Silverman Rule-of-thumb ROT) margins model	0.519 0.000 -0.366 0.357	0.132	0.74	0.0203	0.00042	0.0161	0.977	0.848	-410.85	-408.87	-410.09	0.994	0.98	-6
Parametric margins with parametric copula (PMPC) model [Parametric copulas settings]	r90Clayton copula with Logistic-Logistic margins	NA	0.150	0.58	0.0278	0.00077	0.0228	0.957	0.786	-377.47	-375.50	-376.71	0.989	0.96	-6.6
Parametric margins with nonparametric copula (PMNPC) model	Beta kernel copula with Logistic-Logistic margins model	0.073	0.169	0.42	0.0270	0.00073	0.0227	0.960	0.787	-380.65	-378.68	-379.90	0.990	0.96	3.8

[Semiparametric copulas distribution settings]	Bernstein copula with Logistic-Logistic marginal model	NA	0.188	0.30	0.0278	0.00077	0.0239	0.958	0.776	-377.63	-375.66	-376.87	0.989	0.96	5.6
	Transformation estimator with Logistic-Logistic margins model	0.519 0.000 -0.372 0.363	0.132	0.74	0.0275	0.00075	0.0208	0.959	0.804	-378.83	-376.86	-378.07	0.989	0.96	-5.6
Nonparametric margins parametric copula (NPMPC) model [Semiparametric copula distribution settings]	r90Clayton copula with KDE (GAUSSIAN)-Silverman Rule-of-thumb (ROT) margins model	NA	0.132	0.74	0.0211	0.00044	0.0176	0.975	0.835	-406.71	-404.74	-405.95	0.993	0.98	-6.3
Note: Beta kernel copula density with KDE (GAUSSIAN-Silverman Rule-of-thumb ROT) Marginal (indicated by bold letter with an asterisk) identified as most parsimonious bivariate joint density framework. Majority of GOF test statistics are in favour of the selected bivariate joint density framework for AMRWT and LF events															

Joint distribution framework (Station 2084) (b)	Bivariate joint density frameworks	Estimated Bandwidth	K-S (Kolmogorov-Smirnov)		RMSE (Root Mean Squared Error)	MSE (Mean Squared Error)	MAE (Mean Absolute Error)	NSE (Nash-Sutcliffe Efficiency)	mNSE (modified Nash-Sutcliffe Efficiency)	AIC (Akaike Information Criterion)	BIC (Bayesian Information Criterion)	HQC (Hannan-Quinn Information Criterion)	IA (Index of Agreement)	R2 (Coefficient of Determinations)	PBIAS (Percent Bias)%
			Statistics	p-value											
Nonparametric margins nonparametric copula (NPMNPC) mode [Nonparametric copulas distribution settings]	Beta kernel copula density with KDE (GAUSSIAN)-Silverman Rule-of-thumb (ROT)- KDE (GAUSSIAN)-Unbiased cross-validation (UCV) Scott and Terrell (1987) Margins	0.102	0.130	0.82	0.0201	0.00040	0.0168	0.980	0.852	-357.36	-355.53	-356.67	0.994	0.98	-0.9
	Bernstein copula with KDE (GAUSSIAN)-Silverman Rule-of-	**NA**	**0.108**	**0.94**	**0.0198**	**0.00039**	**0.0165**	**0.981**	**0.855**	**-358.78**	**-356.95**	**-358.10**	**0.995**	**0.98**	**-2.4**

	thumb (ROT)- KDE (GAUSSIAN)- Unbiased cross-validation (ucv) Scott and Terrell (1987) Margins*													
	Transformation estimator with KDE (GAUSSIAN)- Silverman Rule-of-thumb (ROT)- KDE (GAUSSIAN)- Unbiased cross-validation (UCV) Scott and Terrell (1987) Margins	0.544 0.000 -0.289 0.453	0.152	0.66	0.0216	0.00046	0.0179	0.975	0.842	-350.66	-348.833	-349.977	0.994	0.98	-6.8
Parametric margins with parametric copula (PMPC) model [Parametric copulas settings]	r270BB8 copula with Normal-Lognormal margins	NA	0.173	0.48	0.0321	0.00103	0.0275	0.947	0.759	-312.25	-308.598	-310.885	0.994	0.98	-6.8
Parametric margins with nonparametric copula (PMNPC) model [Semiparametric copulas distribution settings]	Beta kernel copula with Normal-Lognormal margins	0.1022	0.108	0.94	0.0270	0.00073	0.0217	0.962	0.810	-330.03	-328.206	-329.35	0.990	0.97	1.7
	Bernstein copula with Normal-Lognormal margins	NA	0.108	0.94	0.0256	0.00065	0.0206	0.966	0.819	-334.87	-333.046	-334.19	0.991	0.97	-0.2
	Transformation estimator with Normal-Lognormal margins	0.534 0.000 -0.278 0.455	0.130	0.82	0.0265	0.00070	0.0220	0.963	0.807	-331.99	-330.168	-331.312	0.991	0.97	-4.6
Nonparametric margins parametric copula (NPMPC) model [Semiparametric copula distribution	r270BB8 copula with KDE (GAUSSIAN)- Silverman Rule-of-thumb (ROT)- KDE (GAUSSIAN)- Unbiased cross-	NA	0.195	0.34	0.0301	0.00090	0.0249	0.953	0.782	-318.47	-314.82	-317.108	0.991	0.97	-4.6

settings]	validation (UCV) Scott and Terrell (1987) Margins														

Note: Bernstein copula with KDE (GAUSSIAN)-Silverman Rule-of-thumb (ROT)- KDE (GAUSSIAN)-Unbiased cross-validation (ucv) Scott and Terrell (1987) Margins (indicated by bold letter with an asterisk) describe most parsimonious joint density structure between AMRWT and LF events for station 2084. Majority of GOF test are in favour of the selected joint density.

Joint distribution framework (Station 2106) (c)	Bivariate joint density frameworks	Estimated bandwidth	K-S (Kolmogorov-Smirnov)		RMSE (Root Mean Squared Error)	MSE (Mean Squared Error)	MAE (Mean Absolute Error)	NSE (Nash-Sutcliffe Efficiency)	mNSE (modified Nash-Sutcliffe Efficiency)	AIC (Akaike Information Criterion)	BIC (Bayesian Information Criterion)	HQC (Hannan-Quinn Information Criterion)	IA (Index of Agreement)	R2 (Coefficient of Determinations)	PBIAS (Percent Bias)%
			Statistics	p-value											
Nonparametric margins nonparametric copula (NPMNPC) mode [Nonparametric copulas distribution settings]	**Beta kernel copula density with KDE (GAUSSIAN)-Unbiased cross-validation (UCV) Scott and Terrell (1987)-KDE (GAUSSIAN)-Sheather and John via (ste-solve the equation) Marginal Model**	0.059	**0.187**	**0.36**	**0.0219**	**0.00048**	**0.0172**	**0.973**	**0.829**	**-364.75**	**-362.87**	**-364.04**	**0.993**	**0.97**	**4.3**
	Bernstein copula with KDE (GAUSSIAN)-Unbiased cross-validation (UCV) Scott and Terrell (1987)-KDE (GAUSSIAN)-Sheather and John via (ste-solve the equation) Marginal Model	NA	0.229	0.16	0.0262	0.00069	0.0210	0.961	0.792	-347.34	-345.47	-346.63	0.990	0.97	9.3
	Transformation estimator with KDE (GAUSSIAN)-Unbiased cross-validation (ucv) Scott and Terrell (1987)-KDE	0.525 0.000 -0.411 0.343	0.166	0.51	0.0220	0.00049	0.0167	0.972	0.834	-364.50	-362.63	-363.80	0.993	0.97	-3.9

	(GAUSSIAN)-Sheather and John via (STE-solve the equation) Marginal Model														
Parametric margins with parametric copula (PMPC) model [Parametric copulas settings]	r90Joe copula with Logistic-Gumbel marginals	NA	0.166	0.51	0.0318	0.00101	0.0244	0.943	0.758	-328.91	-327.03	-328.20	0.985	0.95	-5.5
Parametric margins with nonparametric copula (PMNPC) model [Semiparametric copulas distribution settings]	Beta kernel copula with Logistic-Gumbel marginals	0.059	0.208	0.24	0.0326	0.00106	0.0238	0.940	0.764	-326.45	-324.58	-325.75	0.984	0.94	4
	Bernstein copula with Logistic-Gumbel marginals	NA	0.250	0.09	0.0361	0.00130	0.0268	0.926	0.734	-316.61	-314.74	-315.90	0.980	0.93	7.3
	Transformation estimator with Logistic-Gumbel marginals	0.536 0.000 -0.406 0.343	0.166	0.51	0.0335	0.00112	0.0255	0.937	0.747	-324.03	-322.16	-323.32	0.983	0.94	-3.8
Nonparametric margins parametric copula (NPMPC) model [Semiparametric copula distribution settings]	r90Joe copula with KDE (GAUSSIAN)-Unbiased cross-validation (UCV) Scott and Terrell (1987)-KDE (GAUSSIAN)-Sheather and John via (ste-solve the equation) Marginal Model	NA	0.187	0.36	0.0224	0.00050	0.0175	0.971	0.827	-362.45	-360.58	-361.74	0.992	0.97	-5.6

Note: Beta kernel copula density with KDE (GAUSSIAN)-Unbiased cross-validation (UCV) Scott and Terrell (1987)-KDE (GAUSSIAN)-Sheather and John via (STE-solve the equation) margins Model margins (indicated by bold letter with an asterisk) describe most parsimonious joint density structure for station 2106

Joint distribution framework (Station 2415) (d)	Bivariate models	Estimated bandwidth	K-S (Kolmogorov-Smirnov)		RMSE (Root Mean Squared Error)	MSE (Mean Squared Error)	MAE (Mean Absolute Error)	NSE (Nash-Sutcliffe Efficiency)	mNSE (modified Nash-Sutcliffe Efficiency)	AIC (Akaike Information Criterion)	BIC (Bayesian Information Criterion)	HQC (Hannan-Quinn Information Criterion)	IA (Index of Agreement)	R2 (Coefficient of Determinations)	PBIAS (Percent Bias) %
			Statistics	p-value											
Nonparametric margins nonparametric copula (NPMNPC) mode [Nonparametric copulas distribution settings]	Beta kernel copula density with KDE (GAUSSIAN)-Silverman Rule-of-thumb (ROT) marginals Model*	0.153	0.113	0.02	0.0253	0.00064	0.0196	0.983	0.877	-321.54	-319.76	-320.88	0.995	0.98	-1.3
	Bernstein copula with KDE (GAUSSIAN)-Silverman Rule-of-thumb (ROT) marginals Model	NA	0.091	0.98	0.0256	0.00065	0.0212	0.982	0.867	-320.31	-318.52	-319.65	0.995	0.99	-4.1
	Transformation estimator with KDE (GAUSSIAN)-Unbiased cross-validation (ucv) Scott and Terrell (1987)-KDE (GAUSSIAN)-Sheather and John via (ste-solve the equation) Marginal Model	**0.538 0.000 -0.165 0.504**	**0.090**	**0.99**	**0.0248**	**0.00061**	**0.0209**	**0.984**	**0.880**	**-323.14**	**-321.35**	**-322.48**	**0.996**	**0.99**	**-4.5**
Parametric margins with parametric copula (PMPC) model [Parametric copulas settings]	r90Tawn Type 1 copula with Logistic-GEV marginals	NA	0.204	0.31	0.0311	0.00097	0.0243	0.974	0.848	-301.19	-297.62	-299.86	0.993	0.98	-3.8
Parametric margins with nonparametric copula (PMNPC) model [Semiparametric copulas distribution settings]	Beta kernel copula with Logistic-GEV marginals	0.153	0.092	0.98	0.0302	0.00091	0.0232	0.975	0.855	-305.73	-303.94	-305.07	0.994	0.98	1
	Bernstein copula with Logistic-GEV marginals	NA	0.092	0.98	0.0295	0.00087	0.0226	0.976	0.859	-307.78	-306.00	-307.12	0.994	0.98	-0.1
	Transformation estimator with Logistic-GEV marginals	0.538 0.000 -0.166 0.557	0.113	0.93	0.0289	0.00083	0.0218	0.977	0.863	-309.70	-307.91	-309.04	0.995	0.98	-1.8

Nonparametric margins parametric copula (NPMPC) model [Semiparametric copula distribution settings]	R90 Tawn Type 1 copula with KDE (GAUSSIAN)-Silverman Rule-of-thumb (ROT) marginals Model	NA	0.227	0.20	0.0277	0.00077	0.0241	0.980	0.849	-311.39	-307.82	-310.07	0.995	0.99	-6.8

Note: Transformation estimator with KDE (GAUSSIAN) Unbiased cross-validation (UCV) Scott and Terrell (1987)-KDE (GAUSSIAN) Sheather and John via (ste-solve the equation) margins model * (indicated by bold letter with an asterisk) describe most parsimonious joint density structure for station 2415

Joint distribution framework (Station 2473) (e)	Bivariate models	Estimated bandwidth	K-S (Kolmogorov-Smirnov)		RMSE (Root Mean Squared Error)	MSE (Mean Squared Error)	MAE (Mean Absolute Error)	NSE (Nash-Sutcliffe Efficiency)	mNSE (modified Nash-Sutcliffe Efficiency)	AIC (Akaike Information Criterion)	BIC (Bayesian Information Criterion)	HQC (Hannan-Quinn Information Criterion)	IA (Index of Agreement)	R2 (Coefficient of Determinations)	PBIAS (Percent Bias)%
			Statistics	p-value											
Nonparametric margins nonparametric copula (NPMNPC) mode [Nonparametric copulas distribution settings]	Beta kernel copula density with KDE (GAUSSIAN)-Silverman Rule-of-thumb (ROT) marginals Model	0.096	0.16667	0.6994	0.02376916	0.000564973	0.02002345	0.9665337	0.7968216	-267.234	-265.651	-266.682	0.9918376	0.98	-8.1
	Bernstein copula with with KDE (GAUSSIAN)-Silverman Rule-of-thumb (ROT) marginals Model	NA	0.16667	0.6994	0.02479328	0.0006147066	0.0210321	0.9635877	0.7865869	-264.197	-262.614	-263.644	0.9912055	0.98	-9
	Transformation estimator with KDE (GAUSSIAN)-Unbiased cross-validation (UCV) Scott and Terrell (1987)-KDE (GAUSSIAN)-Sheather and John via (ste-solve the equation) Marginal Model	0.563 0.000 -0.354 0.440	0.19444	0.5041	0.03136543	0.0009837901	0.02749136	0.9417249	0.7210446	-247.268	-245.684	-246.715	0.9856768	0.98	-14.8
Parametric margins with parametric copula (PMPC)	r90 Clayton copula with Logistic-Lognormal	NA	0.25	0.2106	0.03646772	0.001329894	0.0303152	0.9212233	0.6923911	-236.416	-234.832	-235.863	0.9819633	0.96	-12.4

model [Parametric copulas settings]	marginals														
Parametric margins with nonparametric copula (PMNPC) model [Semiparametric copulas distribution settings]	**Beta kernel copula with Logistic--Lognormal marginals model***	0.096	**0.13889**	**0.8782**	**0.02354294**	**0.0005542702**	**0.01948145**	**0.9671677**	**0.8023213**	-267.923	-266.339	-267.37	**0.9923001**	**0.98**	**-4.2**
	Bernstein copula with Logistic-Lognormal marginals	NA	0.13889	0.8782	0.02480537	0.0006153063	0.02060802	0.9635522	0.79089	-264.162	-262.579	-263.609	0.9915322	0.98	-5.1
	Transformation estimator with Logistic--Lognormal marginals	0.558 0.000 -0.352 0.432	0.16667,	0.6994	0.0285316	0.000814052	0.02450298	0.9517794	0.7513678	-254.086	-252.502	-253.533	0.988637	0.98	-11.1
Nonparametric margins parametric copula (NPMPC) model [Semiparametric copula distribution settings]	r90 Clayton copula with KDE (GAUSSIAN)-Silverman Rule-of-thumb (ROT) marginals Model	NA	0.25	0.2106	0.03567015	0.00127236	0.03014981	0.9246314	0.6940692	-238.008	-236.424	-237.455	0.988637	0.97	-16

Note: Beta kernel copula with Logistic--Lognormal marginals (indicated by bold letter with an asterisk) describe most parsimonious joint density structure for station 2473

Table 2: Comparing bivariate joint (both OR- and AND case) versus univariate return periods for (a) station 2044 (b) station 2084 (c) station 2106 (d) station 2415 (e) station 2473

(a)	Station 2044								
RPs (years)	AEP (Annual Exceedance probabilities)	ANEP (Annual Non-Exceedance Probability)	Annual Maximum River Water Temperature (AMRWT) (℃)	Corresponding Low Flow (LF) ($\mathrm{m}^3/\mathrm{sec}$) (Specific discharge ($\frac{\mathrm{m}^3}{\mathrm{sec}}/km^2$))	Joint cumulative distribution function (JCDF)	Univariate Return periods, $\mathrm{T_{RWT}}$ (YEARS)	Univariate Return periods, $\mathrm{T_{LF}}$ (YEARS)	$\mathrm{T^{OR}_{RWT,\,LF}}$ (OR-JRP) (YEARS)	$\mathrm{T^{AND}_{RWT,\,LF}}$ (AND-JRP) (YEARS)
2	0.5	0.5	25.06	14.86 (0.008693)	0.163778	1.99	2.00	1.19	6.10
5	0.2	0.8	26.24	19.09 (0.011168)	0.605622	4.99	5.00	2.53	177.87
10	0.1	0.9	26.82	21.26 (0.012437)	0.801147	9.99	9.99	5.02	870.47
20	0.05	0.95	27.25	23.66 (0.013841)	0.900564	20.00	20	10.05	1775.88
30	0.033333	0.966667	27.45	25.99 (0.015204)	0.933731	30.00	29.99	15.09	2515.09

50	0.02	0.98	27.67	30.09 (0.017602)	0.960254	50.00	49.99	25.16	3937.00
79	0.012658	0.987342	27.84	34.06 (0.019925)	0.974851	79.00	78.99	39.76	5980.86
100	0.01	0.99	27.92	34.82 (0.020369)	0.980134	99.99	99.99	50.33	7457.12

(b) Station 2084

RPs (years)	AEP (Annual Exceedance probabilities)	ANEP (Annual Non-Exceedance Probability)	Annual Maximum River Water Temperature (AMRWT) (℃)	Corresponding Low Flow (LF) ($\mathrm{m^3/sec}$) (Specific discharge ($\frac{\mathrm{m^3}}{\mathrm{sec}}/km^2$)	Joint cumulative distribution function (JCDF)	Univariate Return periods, $\mathrm{T_{RWT}}$ (YEARS)	Univariate Return periods, $\mathrm{T_{LF}}$ (YEARS)	$\mathrm{T^{OR}_{RWT,\,LF}}$ (OR-JRP) (YEARS)	$\mathrm{T^{AND}_{RWT,\,LF}}$ (AND-JRP) (YEARS)
2	0.5	0.5	18.56	8.46 (0.026878)	0.180105	2.00	2.00	1.21	5.55
5	0.2	0.8	20.05	15.15 (0.048132)	0.609638	5.00	5.00	2.56	103.78
10	0.1	0.9	20.89	19.01 (0.060395)	0.801274	9.99	10.00	5.03	784.55
20	0.05	0.95	21.67	25.68 (0.081586)	0.900338	20	20.00	10.03	2962.96
30	0.033333	0.966667	22.03	28.16 (0.089465)	0.933537	30.00	29.99	15.04	4909.18
50	0.02	0.98	22.37	74.43 (0.236466)	0.960119	49.99	50	25.07	8396.30
79	0.012658	0.987342	22.61	76.45 (0.242883)	0.974759	79.00	78.99	39.61	13227.51
100	0.01	0.99	22.73	76.97 (0.244536)	0.98006	99.99	99.99	50.14	16694.49

(c) Station 2106

RPs (years)	AEP (Annual Exceedance probabilities)	ANEP (Annual Non-Exceedance Probability)	Annual Maximum River Water Temperature (AMRWT) (℃)	Corresponding Low Flow (LF) ($\mathrm{m^3/sec}$) (Specific discharge ($\frac{\mathrm{m^3}}{\mathrm{sec}}/km^2$))	Joint cumulative distribution function (JCDF)	Univariate Return periods, $\mathrm{T_{RWT}}$ (YEARS)	Univariate Return periods, $\mathrm{T_{LF}}$ (YEARS)	$\mathrm{T^{OR}_{RWT,\,LF}}$ (OR-JRP) (YEARS)	$\mathrm{T^{AND}_{RWT,\,LF}}$ (AND-JRP) (YEARS)
2	0.5	0.5	21.65	4.31 (0.008693)	0.140403	2.00	2.00	1.16	7.12
5	0.2	0.8	22.66	6.10 (0.011168)	0.606825	5.00	4.99	2.54	146.53
10	0.1	0.9	23.48	7.07 (0.012437)	0.80129	10.00	10	5.03	775.61
20	0.05	0.95	24.13	7.80 (0.013841)	0.900493	20.00	19.99	10.04	2028.80
30	0.033333	0.966667	24.45	8.16 (0.015204)	0.933654	30.00	30.00	15.07	3126.95
50	0.02	0.98	25.00	8.54 (0.017602)	0.96019	50.00	49.99	25.11	5252.10
79	0.012658	0.987342	25.38	8.82 (0.019925)	0.974804	78.99	79.00	39.68	8312.55
100	0.01	0.99	25.49	8.95 (0.020369)	0.980095	99.99	99.99	50.23	10515.25

(d)			Station 2415						
RPs (years)	AEP (Annual Exceedance probabilities)	ANEP (Annual Non-Exceedance Probability)	Annual Maximum River Water Temperature (AMRWT) (℃)	Corresponding Low Flow (LF) ($\mathrm{m}^3/\mathrm{sec}$) (Specific discharge ($\frac{\mathrm{m}^3}{\mathrm{sec}}/km^2$))	Joint cumulative distribution function (JCDF)	Univariate Return periods, $\mathrm{T}_{\mathrm{RWT}}$ (YEARS)	Univariate Return periods, T_{LF} (YEARS)	$\mathrm{T}^{\mathrm{OR}}_{\mathrm{RWT,\,LF}}$ (OR-JRP) (YEARS)	$\mathrm{T}^{\mathrm{AND}}_{\mathrm{RWT,\,LF}}$ (AND-JRP) (YEARS)
2	0.5	0.5	24.65	4.77 (0.011414)	0.216528	2.00	1.99	1.25	4.98
5	0.2	0.8	25.27	6.29 (0.015051)	0.625661	5.00	5.00	2.62	53.85
10	0.1	0.9	25.61	7.73 (0.018496)	0.804999	10.00	9.99	5.05	474.00
20	0.05	0.95	25.95	9.33 (0.022325)	0.901077	20.00	20.00	10.01	5367.68
30	0.033333	0.966667	26.20	10.01 (0.023952)	0.933796	29.99	30.00	15.01	19801.98
50	0.02	0.98	26.60	11.17 (0.026728)	0.960169	50.00	49.99	25.00	78125
79	0.012658	0.987342	26.84	11.71 (0.02802)	0.974756	79.00	78.99	39.50	227272.72
100	0.01	0.99	26.92	11.87 (0.028403)	0.980048	99.99	100.00	50.00	400000

(e)			Station 2473						
RPs (years)	AEP (Annual Exceedance probabilities)	ANEP (Annual Non-Exceedance Probability)	Annual Maximum River Water Temperature (AMRWT) (℃)	Corresponding Low Flow (LF) ($\mathrm{m}^3/\mathrm{sec}$) (Specific discharge ($\frac{\mathrm{m}^3}{\mathrm{sec}}/km^2$))	Joint cumulative distribution function (JCDF)	Univariate Return periods, $\mathrm{T}_{\mathrm{RWT}}$ (YEARS)	Univariate Return periods, T_{LF} (YEARS)	$\mathrm{T}^{\mathrm{OR}}_{\mathrm{RWT,\,LF}}$ (OR-JRP) (YEARS)	$\mathrm{T}^{\mathrm{AND}}_{\mathrm{RWT,\,LF}}$ (AND-JRP) (YEARS)
2	0.5	0.5	16.53	178.26 (0.028299)	0.182559	2.00	2.00	1.22	5.47
5	0.2	0.8	17.44	223.86 (0.035538)	0.611167	5.00	5	2.57	89.55
10	0.1	0.9	17.98	252.16 (0.04003)	0.801489	10.00	9.99	5.03	671.72
20	0.05	0.95	18.47	278.21 (0.044166)	0.900412	20.00	20	10.04	2426.59
30	0.033333	0.966667	18.75	292.82 (0.046485)	0.93358	30.00	30.00	15.05	4065.04
50	0.02	0.98	19.09	310.77 (0.049335)	0.96014	50.00	50	25.08	7127.58
79	0.012658	0.987342	19.40	326.53 (0.051837)	0.974771	79.00	79.00	39.63	11428.57
100	0.01	0.99	19.56	334.56 (0.053112)	0.980069	100	100	50.17	14513.79